

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 13:11:02 ; Search time 4399.1 Seconds

(without alignments)
11520.387 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcataagacaa.....ctttaactagtaactgct 3072

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_btg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_dr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_om:*

20: em_or:*

21: em_ov:*

22: em_pat:*

23: em_ph:*

24: em_pl:*

25: em_ro:*

26: em_sts:*

27: em_sy:*

28: em_un:*

29: em_vi:*

30: em_hgtg_hum:*

31: em_hgtg_inv:*

32: em_hgtg_rod:*

33: em_hgtg_hum:*

34: em_hgtg_inv:*

35: em_hgtg_rod:*

36: em_hgtg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3072	100.0	3133	9	AY032589	AY032589 Homo sapi
2	3067.2	99.8	3219	9	AY035391	AY035391 Homo sapi
3	3065.6	99.8	3396	9	AY027787	AY027787 Homo sapi
4	3065.6	99.8	3581	9	AF376061	AF376061 Homo sapi
5	2001.6	65.2	160583	2	AC010968	AC010968 Homo sapi
6	1998.4	65.1	138909	9	CNS01053	AL121653 BAC seque
7	1316.4	42.9	1335	9	IR2005417	AL389934 Homo sapi
8	815.4	26.5	1395	9	AY027788	AY027788 Homo sapi
9	461	15.0	768	9	AY027789	AY027789 Homo sapi
10	292.2	9.5	185281	2	AC011232	AL121658 BAC seque
11	292	8.8	152692	9	CNS01058	AY027790 Homo sapi
12	269	8.8	578	9	AY027790	G55568 SHGC-100923
13	170	5.5	553	11	G55568	AF242431 Mus muscu
14	68.4	2.2	117791	10	AF242431S1	AF242431 Mus muscu
15	67.8	2.2	4752	10	AF135490	AF135490 Mus muscu
16	67.8	2.2	4815	10	AF102871	AF102871 Mus muscu
17	67.8	2.2	4823	10	AF135489	AF135489 Mus muscu
18	67.8	2.2	5362	10	AF135491	AF135491 Mus muscu
19	67.8	2.2	5497	10	AF007769	AF007769 Mus muscu
20	67.8	2.2	90650	10	AF242431S2	AF242432 Mus muscu
21	66.8	2.2	6829	10	AF135494	AF135494 Mus muscu
22	66.2	2.2	179252	10	AF131205	AF242433 Mus muscu
23	65.2	2.1	41613	10	AF242433S1	AF242435 Mus muscu
24	64.6	2.1	14784	10	AF242433S3	AF135492 Mus muscu
25	62	2.0	5269	10	AF135492	E23944 Excessive o
26	59	1.9	5366	6	E23944	E24990 Apoptosis-1
27	59	1.9	5366	6	E24990	E23943 Excessive o
28	59	1.9	5502	6	A64509	E24989 Apoptosis-1
29	59	1.9	5984	6	E23943	A64529 Sequence 21
30	59	1.9	5984	6	E24989	U19251 Homo sapien
31	59	1.9	6124	6	A64529	A64510 Sequence 2
32	59	1.9	6124	6	HSU19251	A64531 Sequence 23
33	59	1.9	6133	6	A64510	AC005031 Homo sapi
34	59	1.9	6228	9	AC005031	U80017 Homo sapien
35	59	1.9	82976	9	HSU80017	AC010272 Homo sapi
36	59	1.9	131078	9	HSU80017	AC012369 Homo sapi
37	59	1.9	135306	2	AC010272	AC022119 Homo sapi
38	59	1.9	155809	2	AC012369	AC010237 Homo sapi
39	59	1.9	155974	9	AC022119	AC044797 Homo sapi
40	59	1.9	168814	9	AC010237	U21913 Human neuro
41	59	1.9	190871	9	AC044797	AX085744 Sequence
42	56.4	1.8	716	9	HSNA1P1	AC091756 Sus scrof
43	44.2	1.4	1141	6	AX083744	AC091793 Felis cat
44	44	1.4	177639	2	AC091756	
45	42.8	1.4	158082	2	AC091793	

ALIGNMENTS

RESULT	1	ALIGNMENTS
AY032589	AY032589	3133 bp mRNA
LOCUS	AY032589	Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.
DEFINITION	AY032589.1	GI:13899172
ACCESSION	AY032589	
VERSION	AY032589	
KEYWORDS		human.
SOURCE		human.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS		Geddes, B.J., Wang, L., Huang, W.-J., Lavellee, M., Manji, G.A., Brown, M., Jurman, M., Morganstern, J., Merriam, S., Glucksmann, A., Distefano, P.S. and Bertin, J.
TITLE		Human CARD12 is a Novel CED4/Apaf-1 Family Member That Induces Apoptosis
JOURNAL		Biochem. Biophys. Res. Commun. 284 (1), 77-82 (2001)
PUBMED		11374873

REFERENCE 2 (bases 1 to 3133)
AUTHORS Bertin, J
TITLE Direct Submission
JOURNAL Submitted (15-APR-2001) Neurobiology, Millennium Pharmaceuticals
Inc., 640 Memorial Drive, Cambridge, MA 02139, USA
FEATURES
source location/Qualifiers
1..3133
/organism="Homo sapiens"
/db_xref="taxon:9606"
36..3110
/note="CARD12: CED4/paf-1 family member"
/codon_start=1
/product="caspase recruitment domain protein 12"
/protein_id="AAK38730.1"
/db_xref="GI:13899173"
CDS
misc_feature
36..299
/note="CARD; Region: caspase recruitment domain"
misc_feature
532..1406
/note="NBS; Region: nucleotide-binding s.te"
misc_feature
2001..3107
/note="LRR; Region: leucine-rich repeats"
BASE COUNT 903 a 691 c 729 g 810 t
ORIGIN
Query Match 100.0%; Score 3072; DB 9; Length 1333;
Best Local Similarity 100.0%; Pred No 0;
Matches 3072; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaattcataagaacatagccagccctatccaagaatggaatgac:ggtata 60
36 ATGAAATTCATTAAGGACATATAGCCAGCCCTATTCATAAAGATGGAAAGAC:GTTATA 95
61 aagcaaatcaacagatgactatttgtaagatgcttcgaatcgcgagaagataacatc 120
96 AACCAATATCAATATGACATATTTGATGGAATGTTCTGATGCGAAGAGATACATC 155
121 atttgctcgagaagatggaagatgctgctagaaggaatcatccaatgatttggaa 180
156 ATTTGCTGCGAGAGAGGAGAGATGCTGCTAGAGGATCATTCATCATGATTTTAAAA 215
QY 181 aagggctcagaagctgtaacaccttcttaaatcccttaagaagatggaactatc:tcta 240
216 AAGGGTTCAGAGTCCTGTAACCTCTTCTTAATCCCTTAAGAGATGGAATTCCTTA 275
Db 241 ttccaggaacttgatgacaaagtcttttccatcagaacacagaagaagacttgaa:gac 300
276 TTTTCAGGACTTGATGACAAAGCTTTTTCATCAGACATCAGAGAGACTTGGA:GAT 335
QY 301 ttggcctcagaatctaaaggaactgtacacatcccatcttctggaactttatcc:ctc 360
336 TTGGCTCAGATTTAAAGGACTGTGACCATACCCCATCTTTTGAAGCTTTTATCCC:TTT 395
QY 361 ggtgaagaatctgacattattttaaactgaagaacacctcagaagactgtacctg:g 420

Db 396 GGTGAAGATATTGACATATTTTAACTTGAAAGACACCTTCACAGAACCTGCTGTGG 455
QY 421 aggaaggaccacaccacatcacccgctgagagcgtgacccctgaatgacctctgacgct 480
Db 456 AGGAAGACCAACACATCATCCCGCTGAGACGCTGACCTTAATAGGCTCTCTGACAGCT 515
QY 481 ctccagagcccttgatcatctgaagggaattctgcaaaagcgaagtccaactctgtcgag 540
Db 516 CTTCAGAGCCCTGCAATCTTGAAGGGATGTGGCAAGGAGCCATCTGCTGCGAG 575
QY 541 cgaattgcaatgctctgaggctcgaggaagatgcaaggtctgaccaagtccaattctgctc 600
Db 576 CGCATTTGCCATGCTCTGGGGCTCCGAAAGTGTCAAGGCTGTGACCAAGTTCAAAATTCGT 635
QY 601 tctctctccgctcgaagcagggcccaaggatggaacttttgaacctctgatacaactc 660
Db 636 TTTCTTCCTCCGCTCAGCAGGGCCAGGGTGAGCTTTTGAACCTCTGTGATCAACTC 695
QY 661 ctggatataccttgagcaacatcagaagaagcagacatcatgagcatgctctgaagctcgag 720
Db 696 CTGGATATACCTGGCCAAATCAGGAAGCAGACATTCATGCGCATGCTGTGAAGCTGCGG 755
QY 721 cagaaggctcttctctctcttgatgctacaatgaatcaagccccaagaactgccaaga 780
Db 756 CAGAGGCTTCTTCTCTCTGATGCTCAATGATGATCAAGCCCAAGAACGCGCCAGAA 815
QY 781 atcgaagccctgaataaaggaataaacacacgcttcaagaacatgctcgtccaactacc 840
Db 816 ATCGAAGCCCTGATTAAGGAAACACCCGCTTCAAGAACATGCTATCGTCACTACCACTACC 875
QY 841 actgaatgctcgaagacacatcagcgaattgctgctccctgaactggaagtggaggatag 900
Db 876 ACTGATGCTGCTGAGCAGATACGCGCGCTTGGGCTTGAAGCTGCTGAGCTGGGGGATAG 935
QY 901 acgaagaagcagcccaagcgtctcaccgaagaagtctcatcaagaagcttctgaagc 960
Db 936 ACAGAAGACAGGCCCGCAGCTCTCATCCGAGAAGTCTGATCAAGAGAGCTTCTGAGAGC 995
QY 961 ttgttgctcacaatcagaagaatccgaatgcttgaagaatctcaatgaagccctctctt 1020
Db 996 TTGTTCTCCAAATTCAGAAATTCAGAGTCTGAGGAATCTCATGAAGACCCCTCTCTTT 1055
QY 1021 gttggtacactgtgtacatccagatgagatggaagtgaatgacactctacacacaaata 1080
Db 1056 GTGGTATCATCTGTGATCCAGATGGGTGAAGTGAATTCACATCTACACACAAATA 1115
QY 1081 acgctgttccatccctctatgattgtgtgatacagaataaagaacaaacataaagt 1140
Db 1116 AGCTGTTCATACCTCTGATGATCTGTGATACAAATAACAAACAAACATTAAGGT 1175
QY 1141 gttgctgcaagtagacttcaatctggaagcttgagccactgtggaagcctagcttgaagg 1200
Db 1176 GTGGCTGCAAGTACTTTCATCTGGACCTGAGACCTGTGAGAACTTACTCTGAGAGGT 1235
QY 1201 ggttctcccaaatgtgaatttgaaactgcaagatgctgccaagtgtaagaaatgctc 1260
Db 1236 GTGTTCTCCCAAGTTTGAATTCGAAGTGCAGAGATGTCTCCAGCTGATGATGAGATGTC 1295
QY 1261 ctgtgtaacaacttgagctccctctgtaaatatatacagctccaagaagtccaagcacaagtaata 1320
Db 1296 CTGCTGACAACTGGCTCTCTGTAATATATACAGCTCAAAAGGTTCAAGCCAAAGTATTA 1355
QY 1321 tttcttcaagaatctccagagatgacagcagcagaagaactcagcagttatagaag 1380
Db 1356 TTTCTTTCACAAATTCATTCAGAGATACAGAGCGAGGAAGACTTCAGAGTTTATTAAGAG 1415
QY 1381 tctcatgaagcagaagagatgacacaaagggaatggttacttgcagaagaatggttccat 1440
Db 1416 TCTCATGAGCCAGAGAGGAGGACCAAGGGGAATGTTACTTGCAGAAATGAGTTTCCATT 1475
QY 1441 tctggaatcatcatcacttataagcagcctctcggtaaacctgtgtgtaactctgtgga 1500
Db 1476 TCGGACTTATCATCCACTTATACAGAGCTGCTCCGGATACACCTGTGGTCTATGTGGAA 1535


```

/Product="ICE-protease activating factor"
/db_xref="GI:1434215"
/translation="MNPRIKNSRLIORMKTYKQITDDLFVINVIREVNTICE
KVEDDAAGTTHMLIKKSGESDINFLSKWNPPLRDNGO:LFQTSGLDIDA
ODIKDIYPPSPINPBLGEDIITFNKSTFEPYLRKDOH:HFQTSGLDIDA
LOSPCITPESGCKSLIOTAHMWSGKRAIKRFKFEVLI:LSRAOGLFETD
OLDIPGIRKOTPMALIKLORVFLBDQNEKPEFELI:LSRAOGLFETD
VTTPICIRHINOCALUAIEYDMTESAQLIEVILKEIAK:LIIQIKSRILNI
MKTPLPVITTCALDMGESERHSHRTTLTFYDYL:IOKNKH:KVAASDFIRSLD
CGDILQIVSVSHKDFELQDVSYNEVLLTGLCKYQORF:PKYFPHKFOEY
AGRLSILTSHEPEEYTKGNYLQKNWSISDITSTYSLRY:CGSSVATRAVMKH
LAANYQNGELGSLTARKPIKROSLQSVNTEQETLKNINI:SEVECGIHLIOEST
SKASISQEFAPFOGKSLYISGNIPLYLDFEPIHLPNCASALI:FIKLDYFGAMASW
EKAEEDGGIIMEAPETYIPSRVAVSLFPMKQEFRLLEVTLR:FSKLNKODIRYLK
ESRATISRLQIKRCAGVAGSLVSLSTCKNITS:KMEBDAIKL:EGKLNKMLFHL
LSIIDLQORLPGLDLSGLKLNKLT:INDNI:KMEBDAIKL:EGKLNKMLFHL
THLSDISEMDYIVKSLSEPCDLEETQLVSCCISANAVALIAC:NHNLKMLFHL
ENILDERGNEALHELIDRMVLEQULALMPGQDVQGSLSLI:KHLEVPOLFKLGL
KNMRLDPEIRILDAEFGKNPLKNFQDLNLGNKRVSSDGMIAEP:GVFENKLOLVFDFE
STRELPDPLVRLSVLSKLPFLQEARLVGWQFDDDLDSVIT:GAFKLVTA"
Query Match 99.8%; Score 3067.2; DB 9; Length 3219;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Query 1 atgaattcataaaggaacatagcagacccttattcaagaatggaatgactgtatata 60
Db 145 ATGAATTTTCAAAAGGCAATAGCCGAGCCCTATTCAAGAAATGGAATGACGTGTATA 204
QY 61 aagcaaatcacagatgacactattgtatagaaatgttcgaatcgcgaagaatgaatc 120
Db 205 AAGCAAAATCACAGATGACCTATTGTATGAAATGTTCTGAAATCCGAAAGATTA :CATC 264
QY 121 attgtctgcgaagaagtggaagcaagatgcgtcgaaggaatcattccatgattt:aaa 180
Db 265 ATTTGCTGCGGAGAGGTGGAGCAGATGCTGTAGAGGATCATTCATGATTTT :CAAA 324
QY 181 aagggctcagaagtcctgtaacctcttcttaaatcccttaaggagtgaacttc:ctta 240
Db 325 AAGGCTTCAAGATCCTCTTAACCTTTTAAATCCCTTAAAGAGTGAAGTATTC :TCTA 384
QY 241 ttccaaggaactggaatgacaagaatcttttcatcagacatcagaagaagactggt:cgat 300
Db 385 TTTCAGGACTTGAATGAGCAAAAGCTTTTTCATCAGACATCAGAAGAGACTTGG :AGAT 444
QY 301 ttggctcagaattaaagaagctgttaaccatacccatcttctgaacttttcc:icct 360
Db 445 TTGGCTCAGGATTTAAAGACTTGTACATACCCCATCTTTTGAACCTTTTATTC :CCTT 504
QY 361 ggtgaagatattacattattttaacttgaaagaacacttccagaactctgcct:gttg 420
Db 505 GGTGAAGATATTGACATTTATTTTAACTTGAAAGACCTTTCACANAACCTGTGCG :GTGG 564
QY 421 aagaaagacaacaacatcacgcgctgtgagcagcttcaaccttaagtctctgc:ggct 480
Db 565 AGGAAGACCAACACCATACCGCGGTGAGACGCTACCCCTGAATGGCTCTGCG:ggct 624
QY 481 ctccaagagccctcgaacttgaagggaatctggaagaagcaagtccaactctgct:gaag 540
Db 625 CTTCAGAGCCCGTGCATTTTGAAGGGGAATCTGGCAAGGCAAGTCACTGTGCG:GAG 684
QY 541 cgcattgcatactctgtgggctccggaagtgcaaggtctgacaaagtccaattcgtc: 600
Db 685 CGAATTTGCCATCTCTGTGGGCTCCGGAAGGTCAAGGCTCTGTACCAAGTTCAATY :GCTC 744
QY 601 ttcttcctcctctcagaagggcccaaggttgacttttgaaacctctgtgatcaactc 660
Db 745 TTCTTCTCTCCTCTCAGCAGGAGGCCAGGTGACTTTTGAAGAACCTCTGTGATCA :ACTC 804

```

```

QY 661 ctgatataccctgacacatcagaagaacagacattcatagccatgctgctgaagctggg 720
Db 805 CTGATATTACTGGCACAAATCAGAGACGACATTTATGGCCATGCTGCGAAGCTCGG 864
QY 721 caaggggtctcttctctctctgtatgctacaaatgaattcaagccccaagatcccaaga 780
Db 865 CAGAGGTTCTTTCTCTTTGATGGCTACAAATGAATTCAGGCCCAAGACTGCCAGAA 924
QY 781 atcgaagccctcgaataaagaagaacaccgcttcaagaacatgtaactgctacacactacc 840
Db 925 ATCGAACCCCGATTAAGAAACCCAGCCCTTCAAGACACTGGTATCGTACACACTACC 984
QY 841 actgaagctcgaaggaacatagcagagtttggtgacctgactcgaagtggggatag 900
Db 985 ACTGATGCTGAGGACATACGGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044
QY 901 acaagaacagagcccaaggtctcattccagaagaatgatacaagaagcttgctgaagc 960
Db 1045 ACAGAAACAGCGCCCAAGCTCTCATCCGAGAGTCTGATCAAGAGACTTGTGTAAGGC 1104
QY 961 ttgtgtccaaattcaagaatacaaggtcttgaggaaatccatgaagaacctctctt 1020
Db 1105 TTGTGCTCAAAATTCAGAAATCCAGATGCTGTGAGGAATCTCATGAAGACCCCTCTTT 1164
QY 1021 gtgtatcacctgtgcaatcacagatggtgtaagaagtgaattccactctcaacacaaaca 1080
Db 1165 GTGCTATCCTTGTGCAATCCAGATGGGTGAAAGTGAAGTTCCTACTCTCAGACACAAACA 1224
QY 1081 agcgtgttcacatccttctatgatctgtgtgatacagaanaacaaacacataaaggt 1140
Db 1225 ACCCTGTTCATACCTCTTATGATCTGTGTGATACAGAAACAAACACAAATTAAGGT 1284
QY 1141 gtgtgtcgaagtgactcattcgaagctcgaagccactgtggaacctagtctcgaaggt 1200
Db 1285 GTGGCTGCAAGTACTCTTATGATCTGTGTGATACAGAAACAAACACAAATTAAGGT 1344
QY 1201 gtgtctcccaagaatttgattcgaactcgaagatggtgtccagcgttgaatagaatgtc 1260
Db 1345 GTGTTCTCCACAAAGTTTATTTGAAATGCAAGTGGTCCAGCGTGAATGAGGATGTC 1404
QY 1261 ctgtcgaacaactgggtctcctgttaaatatatacagctcacaaggtltcaagcccaagatataa 1320
Db 1405 CTGCTGACACACTGGGCTCTCTGTAAATATATACAGCTCAAAAGTTTCAAGCTTGAAGAA 1464
QY 1321 tctcttacaagatcattccagaagatcacagaagacgaagaactcaagacttattgacg 1380
Db 1465 TTCTTTCAAAAGTATTTCCAGAGATACACAGCAGACGACGAACACTCAGCAGTTTATTGACG 1524
QY 1381 tctcatgagccagagaggtgtgacccaaggggaatggttacttgcagaanaatggttccatt 1440
Db 1525 TCTCATGAGCCCGAGAGAGGTGACCAAGGGGAATGTTACTTGCAGAAATGTTCCATT 1584
QY 1441 tggaaattacatcacttataagcagctgtccgtgtaacccttbtggtlcatctgtgaa 1500
Db 1585 TCGGACATTAATCCACTTATAGCAGCTCTGCGGTACACTGTGGGTCACTGTGGAA 1644
QY 1501 gccacaagagctgtatatagaagcactcgcgaagatgtatacaacgcgctgcctctcga 1560
Db 1645 GCCACCAAGGCTGTATATGAACACCTCCGACAGATGATCAACACGGCTGCTCTCGGA 1704
QY 1561 ctltccatcgcgaagaagccctctcctggaagacaggaacttltgcaaaagtgtgaaanaaaccc 1620
Db 1705 CTTCATCAGCGCAAGAGGCTCTCTGAGACAGGAATCTTTGCAAAATGTGAAANAACCC 1764
QY 1621 actgaagcaagaatttctgaagaagcataaacatcaatccttcttgagggttgcacatccat 1680
Db 1765 ACTGAGCAAGAAATTCGAAGGCCATTAACATCAATTCCTTTGTAGAGTGTGCACTTCAT 1824
QY 1681 ttatatcaagaggtatataccaatcagccctgaagccaagaatttgaagttcttctca 1740
Db 1825 TTATATCAAGAGATTAATCAAAATCAGCCCTAGCCCAAAATTTGAGCTTTCTTTCAA 1884
QY 1741 ggtaaagctatatacaactcagaaggaacatcccgattactatttgacttcttgaa 1800

```

Db 1885 GGTAAAGCTATATATCACTCAGGACATCCCGATTACTATTGACTTCTTGA 1944
 Qy 1801 catttgcccaatttgcgaagtctcttgacctcatctaaacttgacttttaagggaagct 1860
 Db 1945 CATTTGGCCCAATTGGCAAGTCCCTGGACCTTCATTAACTGAGCTTTTATGGGAGCT 2004
 Qy 1861 atgagctctatgggaaaggctcgcagaagacacaggtggaatccactgaaaggagccca 1920
 Db 2005 ATGGCTCTCATGGGAAAAGGCTGCAGACACACAGTGGATCCACATGGAAGAGGCCCA 2064
 Qy 1921 gaaacctacatctccagcagggctglatcttctcttcaacttggaagcaggaaatccag 1980
 Db 2065 GAAACTACATTTCCAGCAGGGGCTGATCTTTGTTCTTCACTGGAGAGGAATTTCAAG 2124
 Qy 1981 actctggaagctcacactccgggaattcagaagttgaaagaagatacacaatctg 2040
 Db 2125 ACTCTGGAGCTCACACTCCGGGATTTTCACCAAGTTGAAATMACAGATATCAGATATCTG 2184
 Qy 2041 gggaataatctagctctgcacaagcctcagcctgcgaataaagagatgctgctgtg 2100
 Db 2185 GGGAAATATTTAGCTCTGCCCCAAGCCTCAGGCTGCATAATAAAGATGCTGGTGTG 2244
 Qy 2101 gctggaaagcctcaagtttggctccctcagcacctgtaagaacattatctctcatggtgaa 2160
 Db 2245 GCTGGAGCCTCAGTTTGGTCTCTCAGACCTGTGAAGACATTTATTTCTCATGTGGGAA 2304
 Qy 2161 gccagctcccctacacatagaagaatgaagacacatcacatctgttaacaaactgaaacc 2220
 Db 2305 GCCACTCCCTCCATAGAAATGATGAGAGCAGACATCATCTGTAAACAACTGAAAAAC 2364
 Qy 2221 ttgagtatctagacctacaagaatcaagcctcggcggtggtctgactacagcttgggt 2280
 Db 2365 TTGAGTATTCATGACCTCAGATCAGACGCTGCGGGGTGCTGCTGACAGCTTGGGT 2424
 Qy 2281 aacttgaagaaccttacaagaagctcaatgtaaaataaagaatgaatgaagaatgct 2340
 Db 2425 AACTTGAAGAACCCTTACAAAGCTCATATGATGATTAACATTAAGATGAATGAAGAAGTCT 2484
 Qy 2341 ataaactagctggaagcctggaagaaacctggaagaagatgcttatttcaatttgaccac 2400
 Db 2485 ATTAACATGCTGAAGGCTGAAACCACTGAAGAAGATGTTGTTTATTCATTGAGCCAC 2544
 Qy 2401 tgcctgacatctggaagggaatggaatcatcatgaatgaatctctctcaagtgaaccctg 2460
 Db 2545 TTGCTGACATTTGGAGAGGGAATGATTCATAGTCAAGTCTCTGTCAAGTCAACCTCT 2604
 Qy 2461 gaccttgaagaatctcaatctagctctcctgctgctgctgctgctgctgctgctgctgctg 2520
 Db 2605 GACCTTGAAGAATTCATTAATGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2664
 Qy 2521 gctcgaatctctcaaatctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2580
 Db 2665 GCTCAGAACTTCACAAATTTGGTCAAACTGACGATTTCTTGAATTCAGAAATTAACCTG 2724
 Qy 2581 gaaaagaatggaatggaagctctctcatgaactgacatgacagagatgaacgtgctagaacag 2640
 Db 2725 GAAAAAGATGGAATGAAAGCTTTCTATGACATGATCGACAGATGACGCTGTAGAACAG 2784
 Qy 2641 ctacacgcactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2700
 Db 2785 CTCACGCGACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2844
 Qy 2701 aacacattggaaggagctcccaactcgtcaagcttgggttgaaaaaacttggaagctcaca 2760
 Db 2845 AAACATTTGGAGAGGCTCCACAACTCGTCAAGCTTGGGTTGAAAAACGGAGACTCACA 2904
 Qy 2761 gatacagaattagaattctagctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2820
 Db 2905 GATACAGAGATTAAGATTTTATGCTGATTTTGTGGAAGAACCTCTGAAAAACTTTCCAG 2964
 Qy 2821 cagttgaatttggcgggaaatcgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2880

Db 2965 CAGTTGAATTTGGCGGGAANCTGTGAGCAGTGAATGATGAGCTTGCCCTCATGGGTGA 3024
 Qy 2881 ttgagaatcttaagcaattgctgttttctgacctagctacaaagaatttctactgac 2940
 Db 3025 TTTGGAATCTTAACCAANTAGTGTGTTTTGACTTATGACTTAAAGAAATTTCTTACTGAT 3084
 Qy 2941 ccagcatagtcagaanaacttagccaagtgtatccaaagttaacttctcgaagaagct 3000
 Db 3085 CCAGCATTAAGTCAGAAACTTAGCCAGGTTATTCCAAGTTACTTTTCTGCAAGAACT 3144
 Qy 3001 aggcctgttggtggcaatttgaatgatgatgctcagtgatgattatcagtgctttaa 3060
 Db 3145 AGCCTGTGTTGGGTGGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3204
 Qy 3061 ctagtaactgct 3072
 Db 3205 CTAGTAACTGCT 3216

RESULT 3
 LOCUS AY027787 3396 bp mRNA PRI 20-JUL-2001
 DEFINITION Homo sapiens CLANA (CLAN1) mRNA, complete cds.
 ACCESSION AY027787
 VERSION AY027787.1 GI:14324112
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 3396)
 AUTHORS Damiano, J.S., Stehlik, C., Pio, F., Godzik, A. and Reed, J.C.
 TITLE Clana, a novel human ced-4-like gene
 JOURNAL Genomics. 75 (1-5), 77-83 (2001)
 MEDLINE 21365712
 PUBMED 11472070
 REFERENCE 2 (bases 1 to 3396)
 AUTHORS Stehlik, C., Damiano, J.S., Pio, F., Godzik, A. and Reed, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
 Research, The Burnham Institute, 10901 North Torrey Pines Road, La
 Jolla, CA 92037, USA

FEATURES
 source
 1..3396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2p22-p21"
 /tissue_type="lung"
 1..3396
 /gene="CLAN1"
 277..3351
 /gene="CLAN1"
 /codon_start=1
 /product="CLANA"
 /protein_id="AAK14776.1"
 /db_xref="GI:14324113"
 Translation="MNFIDKDSRLALIDRMKNTVLIKQTDDLFWNNVLRREVNITCE
 KVEDDAAGIIHMLIKGSESNLEFLSKMWPRLFODLNGOSLFQTSGLDDLA
 QDLKDYHTPEPFLNFPYPLGEDIDILFNLKSTFTPLIMRKQHNHRYEOLILNOLA
 LQSPICIEGSEKSKSTLQRIIDIMLMSGSKALKRKRFVFLRLSAGGLFFETLQD
 QLDLIPGTRKQTFMAMLLKROVLFLLDFGNEFKRONCEITGALIKENRFRNMVY
 VTTTECLRNHRQFGALTAIRYGDMTEDSAQLILREVLIKELABELLIQIOKRLRLN
 MKPLFVVTITAIOMGESERHSHQTLFHTFYDLLLOKNKHNKRGVAASDFISLDH
 RGLDALGEVESHKPEFELQDSVNEEDLLTGLLCKYTAQRFKPKYKFFPKSQEYX
 AGRISSLTSHPEDEVLKNGYLOKMWISLDISTYSSLRLRYCGSSVEATRAVMH
 LAAVYOHCLGLSLIAKRLPRLROESLOSVKYTTQEOILKAININSFVDCGTLIOEST
 SKSLASQEFEPFGKSLYLINGNIPVLPDFEFLPNCASALDFILDFGGMASW
 EKAEEDPGTMEPEAPENYIPSRVSLPENNKQEPRLLEVTLRPSKLNQDRIAYLKL
 IFSSATSLRIDQIKKACAYAGSLSLVLTCKNTYSLAMVBSPLTTEDRHITSTYNLKT
 LSIHDLQNRPLPGLTDSLGNLKNLTKLIMNINMNEEDAKTLEGLKLNKMKCLFLH
 THLSDIEGMQYIVKSLSSPECDLEIQLVSCSLSANAVKILAQNLNLVKSITLDS
 ENYLEKDGNEALHELIDRMNVLEQITLALMLPWGCDVQGSLSILKLHLEVPQLVLCGL

KNMRLLDTEIRILGAFEGKNPLKNFOQLNLAGNRSSDDGLAFWJVENLKOLVFFEDF
STKEFLPDALVRKLSOVLSTKTFLOEARLVGMFDDDLSTVTFKFLVTA"
BASE COUNT 992 a 737 c 793 g 874 t
ORIGIN

Query Match 99.8%; Score 3065.6; DB 9; Length 3336;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3068; Conservative 0; Mismatches 4; Indels 1; Gaps 0;

QY 1 atgaatttcataagaagacatagccgagcccttatccaagaagtgaagtacataata 60
DB 277 ATGAATTTCTAATAAGGACATAGCCGAGCCCTTATTCAAGAAATGGGAATGACTGTTATA 336
QY 61 aagcaaatccagatgacacattctgtatgaatgtctgaatgcgagaagaataa 120
DB 337 AAGCAATCCAGATGACCTATTGTATGSAATGTTCTGAAATGCCGAAGAATNAICATC 336
QY 121 atttgtcgagaagaagtggagagagatgctgctagaagagatcaatcagattt 180
DB 397 ATTGTCTCGAAGAGGTGAGACAGATGCTGTAGAGGAAATCATTCAATGATTTT 456
QY 181 aaggtctcagatgctcgttaaccccttctaatacccttaagaagtgaatactc 240
DB 457 AAGGCTTCAGAGTCTGTATACCTCTTCTTAATCCCTTAAGAGATGGAATATCC 516
QY 241 ttccaagaattgaatggacaagtcttttcaatcaagacatcaagaagaagtggat 300
DB 517 TTTTCAGACTTGAATGACAAAGCTTTTTCATCAGACATCAGAAAGAACTTGG 576
QY 301 ttggctcagaatttaagaagctgttacataccacatcttcttgaacttatac 360
DB 577 TTGGCTCAGAGATTTAAGGACTGTATACCATCCCATCTTTTCTGAACTTTATTC 636
QY 361 ggtgaagataatgaatlatcttaacttgaagaagacacccctcaagaagctgc 420
DB 637 GGTGAAGATATTGACATTTATTTTAACTTGAAAGCACTTTCACAGAACTTATC 696
QY 421 aggaagaagacaacacacacacacgctgtgaagacgtacaccttgaagccctgc 480
DB 697 AGGAAGACCAACACCATCACCGCTGTGAGACACTGACCTTGAAATGGCCCTCG 756
QY 481 ctccaagaagccctgtacatcttaagaaggaatctgccaagaagacatccactgc 540
DB 757 CTTCAGAGCCCTGCATCATTTGAAGGGAATCTGCAAAAGCAAGTCACTCTGCT 816
QY 541 cgaattgcaatgctctggagctcggagaagtgaagtcttgaacaaagtccaatc 600
DB 817 CGCATTTGCCATGCTCTGGGCTCCGGAAGTGCNAAGGCTGTGACCAAGTTCAATTC 876
QY 601 tcttctcctcgtctcagcagagcccaaggttgacatttgaaaacctgtgtacatc 660
DB 877 TTCTTCTCCGCTGTACGACAGGCCCCAGGGTGACTTTTGAACCCCTCTGTGATCA 936
QY 661 ctggatatactgtgacaatcaagaagacagacatctatggccaatgctgtgaagct 720
DB 937 CTGGAAATACCTTGCAACATCAGGAAGCAAGACATTCATGCTGCTGAAGACT 966
QY 721 cagaaggtcttcttcttcttcttgaatgcaatgaattcaagccccaagactccc 780
DB 997 CAGAGGGTCTTTCTCTTGTGATGCTACAAATGAATTCAGCCCAAGAACTCC 1056
QY 781 atcgaagccctgtataagaagaacacgcttcaagaagatgtctcgtcacacac 840
DB 1057 ATCGAAGCCCTGTATTAAGGAACCAACCGCTTCAAGAACTGTCTCATCTCAC 1116
QY 841 acggaatgctcagagacacacagcagattgtgacctgaactcctaagtgaggga 900
DB 1117 ACAGATGCTGAGGACATACGGCAATTTGGTCCCTGACTCTGAGGAGGGAATG 1176
QY 901 acagaagaacagcccaagctctcatccagaagaagtgtcatcaagaagctgtctga 960
DB 1177 ACAGAAAGACAGCCCAAGGCTCTCATCCGAGAGAGTGTGATCAAGAGCTTGTCTGA 1236

QY 961 ttgtctcccaaatccagaatacccaagtgcttgaagaatctcaatgaagaacctctctt 1020
DB 1237 TTGTCTCCCAAAATTCAGAAATCCAGGTGCTTGAGGAATCTCATGAAGAACCCCTCT 1296
QY 1021 ggtgcatcaactgtgcaatccagatgggtgaagtgaagtgtccactctcacacaaca 1080
DB 1297 GTGTCATCACTTGTGCAATCCAGATGGGGAAGGTGAGTTCCACTCTCCACACAAAC 1356
QY 1081 aagctgttccatccctcctatgactcgttgaatcaagaataaacaacaacaataaagt 1140
DB 1357 AGCTGTTCATACCTTTATGATCTGTTGATACAGAAAACAAACCAAACTTAAGGT 1416
QY 1141 gttgcttgaagtgacttcaatcgtgaagccttggaacctgtgaagaactaactcgtga 1200
DB 1417 GTGGCTCCAAAGTACATTTCATTCGAGCCCTTGACACCGTGGAACTAGCTCTGAGG 1476
QY 1201 gtttctcccaagtgttgaatttcgaactgcaagatgtgtccagcgttgaatgagatc 1260
DB 1477 GTGTTCTCCCAAGTTGATTTGGAATCTGCAAGATGTCTCAGCTGATGAGATGTC 1536
QY 1261 gttgtgaacactgtggtcctctgttaatatagactccaaggttccaagccaagtataa 1320
DB 1537 CTGCTGACAACTGGCTCTCTGTAAATATACGTCGAAAGGTTCAAGCCAAAGTATAA 1596
QY 1321 tctttcaagaatcattccaagaagtaacaagcaagaagcaagactcaagcagtttatgac 1380
DB 1597 TTCTTTCACAAATTCATTCAGAGATACACAGCAAGAGAACTCAGCAAGTTTATGAG 1656
QY 1381 tccatgaagccagaagagtgagaccaaagggaatcgttacttgcagaagaatggttccat 1440
DB 1657 TCTCATGAGCCAGAGAGGAGGACCAAGGGGAATGTACTTGCAGAAATGTGTTCCATT 1716
QY 1441 tctgaatcattacatccacttgaagacccctgcagtgtaacccctgtgtgtaactgtgga 1500
DB 1717 TCGACATTTACATCCACTTATATACACCTGCTCCGTAACCTGTGGTGCATCTGTGAA 1776
QY 1501 ggcacagaggtgttgaagaagccctgcagagagatatacaagaagcctgcctctcga 1560
DB 1777 GCCACCAAGGCTTTATGAAGCACTGCAAGCACTGATATCAACACGGCTGCCCTCG 1836
QY 1561 ctctccatgcgaagaagcctctctgagaagaagaaacttctgaagaagtgtgaagaac 1620
DB 1837 CTTTCCATCCCAAGAGGCTCTCTGTGAGCAAGAAATCTTTGCAAGTGTGAHAACAC 1896
QY 1621 actgaagcaagaatcttgaagaagcaatacaatctccttgtagaagtgtgcatcat 1680
DB 1897 ACTGAGCAAGAAATCTGAGAAAGCAATTAATCAATTCCTTGTGAGAGTGTGCAATC 1956
QY 1681 ttatatcaagaagatacatccaaatcagccctgaagcgaagaattgaagcttcttcaa 1740
DB 1957 TTATATCAAGAGATGATCATTCACAAATCAGCCCTGAGCAAGAAATTTAAGCTTTT 2016
QY 1741 ggtataaagcttatataatcaactcaaggaacatcccgatctacttacttcttga 1800
DB 2017 GTTAAAGCTTATATATCAACTCAGGAACATCCCGATTAATTAATTAATTAATTA 2076
QY 1801 catctgcccaatgtgcaagtgtctcgtgactcatttaactgaacttataagtgaggat 1860
DB 2077 CATTTGCCCAATGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTG 2136
QY 1861 atgtgtcattgaagaagaagctgtcagaagaacagagtggaatccaatgaagaagagccca 1920
DB 2137 ATGGCTTCAATGGGAAAAGGCTGCAGAAAGCACAGGTGGAATCCACATGGAAGAGCCCA 2196
QY 1921 gaactacatctcccaagcagaggtgtatcttgttcttcaacttgaagaagaagaaatc 1980
DB 2197 GAACACTACATTCACAGCAGGCTGTATCTTGTCTTCAACTGGAAGCAAGGATCAG 2256
QY 1981 actctggaagtgcaactccggagattcaagaagaatgaataagaagataatcatctgt 2040
DB 2257 ACTGTGAGGTACACTCCGGGATTTTACAGCAAGTTGAATTAAGCAAGATATCATGAT 2316

Oy 2041 gggaaataatcagctctcccaagaagcctcgaagcctgaataaagaatgctgctg 2100
 |||||
 Db 2317 GGGAAATATTCAGCTCTGCCCAAGCCTCAGGCTGCAATTAAGATGTCTGTGTG 2376
 |||||
 Oy 2101 gctggaagcctcagcttctgctccagcagcctgtaagaacattatctctcagtg 2160
 |||||
 Db 2377 GCTGGAACCTCAGTGTGTGCTCTCAGCACCTGTAGAACAATTATTTCTCTCATG 2436
 |||||
 Oy 2161 gccagctcccccacacataagaatgagaggaacatcaatctgttaacaacctgaaac 2220
 |||||
 Db 2437 GCCACTCCCTCACCATTGAGAGATGAGGCAACATCATCTGTAAACAACCTGAAACC 2496
 |||||
 Oy 2221 ttgaattatcattgacctcaagaatcaacagcgtccgggtgtgtctgacctgagct 2280
 |||||
 Db 2497 TTGAGTATTCATGACCTTACAGAAATTAAGGCTGCCGGGTGCTTACTGACAGCTT 2556
 |||||
 Oy 2281 aacttgaagaacctatacaagctcaataatgatacaataaagaatgaaatgct 2340
 |||||
 Db 2557 AACTTGAAGAACCCTTACAAAGCTCATATGATTAACATTAAGATGAATGAAGAAGAT 2616
 |||||
 Oy 2341 ataaactagctgaaagcctgaaacaaacctgaaagaatgctgttattctcattgac 2400
 |||||
 Db 2617 ATAAAAACAGCTGGAAGGCTGAAAAACCTGAAGAGATGTGTATTTTCATTTGACCC 2676
 |||||
 Oy 2401 ttgtctgacattgagagaggaatggaattacatagctcaagctctgtaagtaacct 2460
 |||||
 Db 2677 TTGCTGTGACATTTGGAGAGGGAATGATTAACATAGTCAAGTCTCTCTCAAGTGA 2736
 |||||
 Oy 2461 gaccttgaagaatcaatctagctctcctgctgctgctgctgcaaatgcaatgca 2520
 |||||
 Db 2737 GACCTTGAAGAATTCATTAATGCTGCTGCTGCTGCTGCTGCAAAATGCAATGCAAT 2796
 |||||
 Oy 2521 gctcagaatcttcacaaattgctgtaaacctgagcattcttgattatcagaat 2580
 |||||
 Db 2797 GCTCAGAACTCTTCAAAATTTGGTCAAACTGAGCAATCTTGTATTCAGAAATTA 2856
 |||||
 Oy 2581 gaaaaagaatgaaatgaaagcctctcactgaactgacagagatgaagctgtaaga 2640
 |||||
 Db 2857 GAAAAAGATGGAATGAAGCTCTTATATCACTGATGACAGAGATGAACGTGCTAGA 2916
 |||||
 Oy 2641 ctcaacgcagatgctgctcctgctgctgctgctgctgctgctgctgctgctgct 2700
 |||||
 Db 2917 CTCACGCGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2976
 |||||
 Oy 2701 aaacatttggagagagctcccaaacctgctcaagctggttggaaacctggaact 2760
 |||||
 Db 2977 AAACATTTTGGAGAGAGCTCCCAACCTGCTCAAGCTTGGGTTGAAAAACTGGA 3036
 |||||
 Oy 2761 gatacagaagatagaattttgagctgcttcttggaaagaacctctgaaaacctcag 2820
 |||||
 Db 3037 GATACAGAGATTAGAAATTTTGGTGCATTTTGGAAAGAACCTCTGAAAAAATCT 3096
 |||||
 Oy 2821 cagttgaaatttggcgggaacatgctgctgagcagtgatgagctgctgctcattg 2880
 |||||
 Db 3097 CAGTTGAATTTGGCGGGAATCGTGTGAGCAGTATGATGGCTTGCCTTCAATGG 3156
 |||||
 Oy 2881 ttgagaatcttaagaactgctgttcttgaacttagtaactaaagaattcttactg 2940
 |||||
 Db 3157 TTTGGAATCTTAACCAATTAAGTGTGTGTGCTTGAATTAAGAAATTTCTACTG 3216
 |||||
 Oy 2941 ccagatattgacagaagaaccttagccaagtgtatccaagtaaccttctcagaagct 3000
 |||||
 Db 3217 CCAGATTATGTCAGAAAACTTAAGCAAGTGTATTCAGATTACTTTCTGCAAGAA 3276
 |||||
 Oy 3001 aggccttctggtggaacttctgagatgagatgctcagctgtatctacagctgct 3060
 |||||
 Db 3277 AGGCTTGTGGGTGCAATTTGATGATGATGATGATGATGATGATGATGATGATG 3336
 |||||
 Oy 3061 ctgagtaactgct 3072
 |||||
 Db 3337 CTAGTAAGTCTGCT 3348
 |||||

RESULT 4
 AF376061
 LOCUS AF376061 3581 bp mRNA PRI 15-MAY-2001
 DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.
 ACCESSION AF376061
 VERSION AF376061.1 GI:14040074
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 3581)
 AUTHORS Gingras, M., Qiu, J. and Margolin, J.F.
 TITLE Differential expression of the caspase recruitment domain protein 12 (CARD12) during monocytic differentiation
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3581)
 AUTHORS Gingras, M., Qiu, J. and Margolin, J.F.
 TITLE Direct Submission
 JOURNAL Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center, Baylor College of Medicine, 6621 Fannin St., MC3-3320, Houston, TX 77030, USA
 FEATURES
 source location/Qualifiers
 1. 3581
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2p21-22"
 /cell_line="U937"
 /cell_type="peripheral blood-derived CD14 mature monocyte"
 490..3564
 /note="CARD12"
 /codon_start=1
 /product="caspase recruitment domain protein 12"
 /protein_id="AAK53443.1"
 /db_xref="GI:14040075"
 /translation="MNFIRKNSRALLDORMGVYIKQTDLPFVNVLNREVNITCE
 KVEDDANRGIITHLNLRKSESNCIFLSLEKEMTPYLPYDNLNGSLFPHQTSGLDDIA
 QDLKDIHTPSFLNFYPLGEDIIDITENLKSTFETPRVLMRDHNRVLEQLGLDA
 LOSPCLIEGSGGKSTLLQRIAMLMGSCCKALTKFKEVFLRLSAOGLFETLGD
 QLLDIPETIKQTFEMALIKLRQVLELDGVNEFKQNPETALIKHRRKNMYI
 VTTTECLRHIRHIGALTAEVGMTEDSAALIREVILKRLAGLILLOKSLRLNL
 MKPTPLVYITCALIQMGSESEPHSQTLFTFTFDLLQKKKKHKYAAADPTRSIDH
 CGDLALGVFSHKFDELDQVSSVNEBALITGLCKYTAQRKPKYKFKHSQET
 AGRLLSLTSHPEEVTKNGVLOKWSISDTSTYSLLRTGSSVPAATAVMKH
 LAAYVQHCCLGLSIAKRLPMROESLOSVSNTTQETLKAININSEFEGILHYOEST
 SKSLISOEPFAPFOGKSLYINSNIPYLPDFEHLPPNCASALDFIKIDYGGAMSG
 EKAEDDGGIHMEDEAPETYPISRAVSLFFMKQGFRTLEVTLDLFSKLNKODIRLYGX
 IFSATSRLRIQIKRCAVAGSLSLVSTCKNITSLEVEASPLTIEDRHITSTYNLKT
 LSLHLDQNLRLPGGLDLSLGNLNLTKLIMDNTKMNEEDIKLAEGIKNLKRCFLHL
 TSLSDIEGMDYIVKLSLSEPCDLEELIOLVSCCLSAVAVKILANLHNLVKSLIDLS
 ENYLEKNGELHELIDIRNMLVQLRLALMLPMGCDVQGSLSLKLHLEKPOLVKGL
 KNNRLDTEIRIIGAFGKPNLKNFOOLNAGRVSDGMLAMGVENLKKLVFFDF
 STKEFLPDPALVRLKLSQVLSKLIPLQEARLVQNPDDDLSTVTGAFKLVTA"
 502..741
 /note="Region: caspase recruitment domain"
 BASE COUNT 1033 a 781 c 843 g 924 t
 ORIGIN
 Query Match 99.8%; Score 3065.6; DB 9; Length 3581;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3068; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 atgaattcataaagaagaatagcagcccttattcaagaagaaggaaatgactgtata 60
 |||||
 Db 490 ATGAATTTCAATGAAGAGCAATAGCCGAGCCCTTATTCAAGAAATGGGAAATGACT 549
 |||||
 Oy 61 aagcaaatcacagatgacctattgtatggaatgcttcaaatcggaagaatcaaacatc 120
 |||||
 Db 550 AAGCAAAATCACAGATGACCTATTGTATGCAATGTTGTAATCCGAAGAAGTAACATC 609
 |||||


```
Db 2770 AACTTGAAGAACTTACAAAGCTCATTAATGATTAACATTAAGATGAATGAAGATGCT 2829
Oy 2341 aataaactagctgaagcctgaataaactgaagaagatgtttattcttaaccac 2400
Db 2830 ATTAAGTACTGAGAGGCTTAAAGCTGAAGAGATGTATTATTCTTATTTGACCCAC 2889
Oy 2401 ttgtctgacatttgagagaggaatgatatacaatgaatcgtctctgcaagtgaacccgt 2460
Db 2890 TTGTCTGACATTGGAGAGGGAATGATTAACATTAAGTCAAGTCTGTCAAGTGAAGACCTGT 2949
Oy 2461 gaccttgaagaatcaatgaatgactctctgctgtctgtctgcaatgcagtgaaatcccta 2520
Db 2950 GACCTTGAAGAAATTCATTAATGATGTCTCCGCTGCTTGTCTGCAAAATGAGTGAATACCTTA 3009
Oy 2521 gctagaactctcaaatcttgtaactgagcaatcttctgattatcagaataattactg 2580
Db 3010 GCTGCAATCTTCACAAATTTGTCAAACTGAGCATTCTTGAATTAACGAAATTAACCTG 3069
Oy 2581 gaaaaagatggaatgaagcctctcatgaactgaatcgaagaatgaacgtgctagaacag 2640
Db 3070 GAAAAAGATGCAATGAAGCTCTTCATGAAGTCAAGTCAAGTGAAGTCAAGTCAAGTCAAG 3129
Oy 2641 ctcaacgcagctgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2700
Db 3130 CTCACCCGACTGATGCTGCTCCCTGGGGGCTGTGACGTGCAAGCAGCAGCTGACGACCTGTTG 3189
Oy 2701 aaaaattggaggaggtgcccaactgctcaagcttggtgaaactggaagactca 2760
Db 3190 AAACATTTGGAGAGGCTCCCAAACTGCTCAAGCTGGGTTGAAAAGTGAAGCTGACCA 3249
Oy 2761 gatacagagattagaatlttagtgcattttctgaaagaacctctgaaactccag 2820
Db 3250 GATACAGAGATTAGATTTTAGTGATTTTGTGAAAGAACCCCTGAAAAGTCTCCAG 3309
Oy 2821 cagttgaacttgcgcgggaaatcgtgtgagcagtgatgagtgctgctcactgagtgta 2880
Db 3310 CAGTTGATTTGGCGGGGAAATCGTGTGAGCAGTATGATGCTGCTGCTGCTGCTGCTGCTG 3369
Oy 2881 ttgaagaacttaagcaatgattgttttttgaacttaagtaagtaattctactcgat 2940
Db 3370 TTTGAGAACTTTAAGCAATTAAGTGTGTTTGTGACTTAAAGAAATTTCTACCTGAT 3429
Oy 2941 ccagcatagtcagagaacttaagcgaagtgtatcgaagtaactttctcagaagact 3000
Db 3430 CCACCATTAAGTCAAGAAACTTACGCAAGTGTATCCAAAGTATTTCTGCAAGAGCT 3489
Oy 3001 aggcctgtgtgggagcaatttgatgatagtatcctcagtgatcattacaggtgtcttaaa 3060
Db 3490 AGGCTTGTGGGTGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3549
Oy 3061 ctgactaactgct 3072
Db 3550 CTAGTAAGTCTGCT 3561
```

```
RESULT 5
AC010968 160583 bp DNA HTG 18-AUG-2000
LOCUS Homo sapiens chromosome 2 clone RP11-9302, WORKING DRAFT SEQUENCE,
DEFINITION 11 unordered pieces.
AC010968
AC010968 5 GI:9845170
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 160583)
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160583)
AUTHORS Waterston, R. H.
```

```
*****
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington
MO 63108, USA University School of Medicine, 4444 Forest Park Parkway, St. Louis,
On Aug 18, 2000 this sequence version replaced gi:8439959.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0093002
----- Summary Statistics -----
Sequencing vector: M13: 59%
Sequencing vector: plasmid: 41%
Chemistry: Dye-Primer ET: 47% of reads
Chemistry: Dye-terminator Big Dye: 53% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 159583; sum-of-Configs
Quality coverage: 6.64 in Q20 bases; agarose-fp
Quality coverage: 6.38 in Q20 bases; sum-of-Configs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1080: contig of 1080 bp in length
* 1081 1180: gap of unknown length
* 1181 2476: contig of 1296 bp in length
* 2477 2576: gap of unknown length
* 2577 5035: contig of 2459 bp in length
* 5036 5135: gap of unknown length
* 5136 9606: contig of 4471 bp in length
* 9607 9706: gap of unknown length
* 9707 17078: contig of 7372 bp in length
* 17079 17178: gap of unknown length
* 17179 27158: contig of 9980 bp in length
* 27159 27258: gap of unknown length
* 27259 45137: contig of 17879 bp in length
* 45138 45237: gap of unknown length
* 45238 65522: contig of 20285 bp in length
* 65523 65623: gap of unknown length
* 65623 91498: contig of 25876 bp in length
* 91499 91598: gap of unknown length
* 91599 116835: contig of 25237 bp in length
* 116836 116935: gap of unknown length
* 116936 160583: contig of 43648 bp in length.
*
* Location/Qualifiers
* 1. 160583
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="2"
* /clone="RP11-9302"
* 1. 1080
* /note="assembly_name:Contig9"
* 1181. 2476
* /note="assembly_name:Contig12"
* 2577. 5035
* /note="assembly_name:Contig13"
* 5136. 9606
* /note="assembly_name:Contig14"
* 9707. 17078
* /note="assembly_name:Contig15"
* 17179. 27158
* /note="assembly_name:Contig16"
```


QY	2121	cctgcagcaccgtaagaacattatctctcaatggttgaaagccagtcctccctacataga	2180
Db	51046	CCTCAGCACCTGTAAAGAACATTTATTTCTCTCATGTGGTGAAGCCACTCCCTCCACCATAGA	51105
QY	2181	agatgaaggagccaccatcatctgtatacaaaccttgaaacctgtgattatcaagacctga	2240
Db	51106	AGATGAGAGGCACATCACTCTGTACAAACCTGTAAACCTGTAAGCTATTTCATGACCTTCA	51165
QY	2241	gaatcaagcgtctcgggtggtctgactgacgacgtctgggtaacctgaaacctcaaca	2300
Db	51166	GAATCAAGCGGTCCGGCGGTATTGTGAATATCATAGTGGTTTGTCCTTTGTTCACTTAAAAA	51225
QY	2301	gctcaataatgataacataaagaatgaatgaagaagatgctataaact	2348
Db	51226	AAATFACAGTGTATTAATTTGGAAAAGAGAGAGAGACTTATTCTT	51273
RESULT	6		
NCBI	501053/c		
DEFINITION	CNS01D53	138909 bp	DNA
KEYWORDS			PRI
SOURCE			BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 164M19
ACCESSION	AL121653	2	GI:7159616
VERSION	AL121653	2	GI:7159616
KEYWORDS	HTG: HTGS_DRAFT; SPG4	genomic DNA	interval.
ORGANISM	human.		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
AUTHORS			
COMMENT			
FEATURES			
source			
BASE COUNT	39243 a	28424 c	29121 g
ORIGIN			
Query Match	65.1%	Score 1998.4	DB 9: Length 138909;
Best Local Similarity	97.3%	Pred. No. 0	
Matches 2032:	Conservative	0: Mismatches	56: Indels
			0: Gaps
QY	261	aagtcctttcatcagacatcagaagaagagacttggacgatttggctcagatlttaaaga	320
Db	75311	AGGCTCTTTTCATCATCAGACATCAGAAAGAGACTTGGACATTTGGCTCAGAGATTTAAAGA	75252
QY	321	cttgcacacatccccatctttcttgcgaacttttatccctctggttgaagatatgacatat	380
Db	75251	CTTGACATCATCCCCATCTTTTCGAACTTTTATCCCTTGATGAAGATATTGACATTTAT	75192
QY	381	ttttaacttgaagaagacacttcaacgaagacgctcctctgtaggaagagacaacacatca	440
Db	75191	TTTTAACTTTGAAAGACACCTTACACAGAACCTGTCTGTGTGAGAGAGACCAACACATCA	75132
QY	441	ccgcgttgagagagctgacccctgaatgagcctcctcctggaagctcttcagagccctgcatat	500

Db	75131	CCGCGTGAAGACGCTGACCCCTGGAATG6CCTCCMGAG6GCTCTTTCAGAG6CCCTGCATCAT	75072
OY	501	tgaaagggaatctggcgaaggcaagtccaactctgtctgcagcgcatctgcgaactctctggg	560
Db	75071	TGAAGGGGAATGTGGCAAAAGGCAAGTCCACTCTGCTGCAGCGCAATTTGCCATGCTTGG66	75012
OY	561	ctcggaaagtgcaaggctctgcaccaagtccaattctctctctctcctcgcgtctcaagag	620
Db	75011	CTCCGGAAGTGCAGAGGGCTGTACCAAGTTCAAATTTGGTCTTTCCTGCTCAGCAG	74952
OY	621	ggcccaaggttgactcttttgaaaacctctgtatccaactctgtalataacttgcacaat	680
Db	74951	GGCCAGGGGTGACCTTTTGTGAACCCCTCTGTATCACTCCGGATTAACCTGGCACAAT	74892
OY	681	caggagaagagacattcatatggccaactctctgaagctgsgaagggtctcttctctct	740
Db	74891	CAGGAAGCAGACATTCAATG6CCTATGCTCTGTGAAGCTGG6GAGAG6GTTCCTTCTCT	74832
OY	741	tgtatgtctacaaatgtaattccaagccccagaagctgcgccaagaatacgaagctctalataaga	800
Db	74831	TGATGGTACATATGAATTCAACGCCCCAGACACTGCCAGAAATCGAAGCCGTGATTAAGA	74772
OY	801	aaaccaagcgtctcaagaacatgtgtcatcgtlccaactaccactgagtgctgtaggacat	860
Db	74771	AAACCAAGCGCTTCAAGAACATG6TGTATGTCACCACTACCACTGTAGTGCCTGAGCACAAT	74712
OY	861	acggaagtttggctggcccgacgtcgtcgaagtgagggaatactgaagaagaagagcccaagc	920
Db	74711	-ACGGCAGTTTGGTGGCCTGTACGCTGAG6GTGG6GATATGACAGAAAGACAG6GCCACAGC	74652
OY	921	tctcactccgaagaagtgctgatalccaagaagagcttgcataagagctgtgtctccaatatcaga	980
Db	74651	TCTCATCCGAGAAAGTGTGATCAAGAGACTTGTCTGAAG6CTTGTTGCTTCAAATTCAGAA	74592
OY	981	atccaagtgtcttgaggaatctcaatgaagaacctctcttgtgtgltcatcaacttgtacaat	1040
Db	74591	ATCCAGGTGCTTGAAGAAATCTCATGAAGAACCCCTCTTGTGGTCATCACTTGTGCATAT	74532
OY	1041	ccaagtgggtgaagaatgggttccactctcaacacaaacaaagcgtctccatctactcta	1100
Db	74531	CCAGATGG6GTGAATGTGATGTTCACTCTTCACACACAAACACAGCGATTCATCACTTCTGA	74472
OY	1101	tgatctgtctgatalacgaagaaacacacacacataaagtgtygtctgataagtygactcat	1160
Db	74471	TGATGTGTTGATACAGCAAAACAAACAAACATTAAGGTGTGGCTGCAAGTGAATTCAT	74412
OY	1161	tcgagagcttggaaccaactgttggaacctagctcttgaaggtygtctccccaagaatttga	1220
Db	74411	TCGAGAGCTTGAGACCTGTGGAGACCTGTGACCTGTGAGGGTGTGTTCTCCCAAAATTGA	74352
OY	1221	tttgcgaacttgcagaatgtgtgtccaagtgtaatgaagatgtctctgtctgcacaacttgggtcct	1280
Db	74351	TTTCCAACTGTCAGAGATGTCCTCCAGGTTAAATGAGATGTCTCGTGAACACGGGCTCCT	74292
OY	1281	ctgtgaataatcaagaatcaaaagttcaagccaaagataaattctcttccaagttacttca	1340
Db	74291	CTGTAAATATACAGCTCAAAGSTTCAACGCCAAAGATATAAATTTCTTTCACAAATCTTCCA	74232
OY	1341	ggagatcacagcagaagcagaagactcagaagttatatttagtcgtlctcaatgcgaagagaggt	1400
Db	74231	GGAGATCACACAGCAGCAAGACTCAGCAGTTTATTGACGTCTCATGTGAGGCCAGAGAGGT	74172
OY	1401	gaccagaagggaatgtgtactctgtcagaanaatgtttccaattccgaacatatactcaactta	1460
Db	74171	GACCAAGGGGAATGTTACTTTCGACAATAATGTTTCCATTTTCGGCATTACATTCACATTA	74112
OY	1461	ttagagactgtctccggatacaacttgggtgtactctgtgaagccaccagaggtcgttatga	1520
Db	74111	TAGCAGCTCTGCTCCGGTACACTGTGGGTCAATCTGTGAAGCCACCAAG6GCTGTTATGAA	74052
OY	1521	gcaacctgcagcagatgtatatacaacagcgtctgcttctcgyaacttccaatcgcgcaagaagcc	1580
Db	74051	GCACCTGCACAGTGTATCAACAGGCTGCTTCTCGACTTTCATTCGCAAGAGGCTC	73992

[illegible]

QY	2115	tttgcctccagcaacccgttaagaacattttctctcatggttggaagccagctccctac	217
Db	361	TTTGCTCTCAGCACCCGTGAAGACATTATTTCTTCATGCTGTGAAGCCAGTCCTCCATC	420
QY	2175	catagaagatgatgagggacacatccatcctgttaacaaaccttgaaaccttgatattcatga	2234
Db	421	CATGAGAGATGAGAGGACCATTCATCTGTAACAAACCTGGAACCTTGAGTATTCATGA	480
QY	2235	cctacgaatccaacagctctgcgggtgtgtctgactgacagcttgggttaacttgaagaacct	2294
Db	481	CCGTACAAATTCACAGCGCTGCCGGGTGGTCTGACTGCACACTTGGGTAACTTGAACAACT	540
QY	2295	tacaagctcctaaatggaataacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaat	2354
Db	541	TACAAACCTCATANTGATGATTAACATAAAGATGATGAAGAAGATGCTATTAATCACTGCTGA	600
QY	2355	aggccttgaaaaacctggaagaagaatgtgtattttcaatttgacccaacttgctccagatctgg	2414
Db	601	AGGCGCTGA AAAACCTGGAAGAAGATGTGTATTTATTTCAATTTGACCACTTGCTGACATTGG	660
QY	2415	agaggggaatgatattacatagtaacatctctgtcaagtgaacctgtgaccttgaagaat	2474
Db	661	AGAGGGGAATGGAATTAACATAACTCAAGCTCTGCTCAACTGGAACCTGTGACCTTGAAGAAAT	720
QY	2475	tcaattgctctccgtcgcgtctgtctcctcaaatgcaatggaataacctctagctccagaattctta	2534
Db	721	TCAATTAAGTCTCTCGCTGCTGCTTCTCTCAANTGCAGTGAATAATCTTAGCTCAAAATCTTCA	780
QY	2535	caatttgatcaaaactgagcaattcttgattatcatcaagaataatcaacttgaaaaaaatgagaa	2594
Db	781	CAATTTGCTCAAACTGAGCATCTCTGATTTATCAGAAATTAACCTGGA AAAAGATGGA	840
QY	2595	tgaagctcttccatggaactgtgtgacaggaatggaacgtgtcctagaagaagctcaacccagctgat	2654
Db	841	TGAAGCTCTTTCATGAATGTGATGCACAGGATGAACGCTGTGAGAACAGCTCACCGCACTGAT	900
QY	2655	gctgcctctggggctgtgacgtgcaagagcagccctgacagcctgtgttgaacaatttgaagaa	2714
Db	901	GCTGCCTGGGGCTGTGACGTGCAGAGGCAACCTGTGACACCTGTTGTAACATTGTGAGAGA	960
QY	2715	ggtccccaacactcgtcaagcttgggttgaaaaaacttgaagactcacagatacagaagattag	2774
Db	961	GGTCCCAACACTCGTCAAGCTTGGGTTGAAAAAAGTGAACATCAGATACAGAGATATAG	1020
QY	2775	aattttagtgatcttttttggaaagaacctctgaaaaacttccagcgagttgaattggc	2834
Db	1021	AATTTTAGGTGATTTTTTGTGAAAAAGAACCTCGA AAAAATTCACAGCACTTGAATTTGGC	1080
QY	2835	gggaaaacgtgtgagcagtgatgaatggccttgcctcatcgggttgatatttgaagaactctaa	2894
Db	1081	GGGAAATCGTGTAGCAGTATGATGATGGCTGCTTCATAGGGGTGATTTTGAACATCTTAA	1140
QY	2895	gcaatgagtgltttttgactttagtaactaaagaatcttcaactgataccagcatatgacag	2954
Db	1141	GCAATTTAGTGTTTTTTGTACTTTAGTACTTAAGAATTTTCACTGTGATCCAGCATTTAGTGTAG	1200
QY	2955	aaaacttagccaagtgatataccaagttaactttctgcaagaagcttagctgtttgggtg	3014
Db	1201	AAAACTTAGCCAACTGTATTATCCAAAGTTAACTTTTTCGCAAGAAAGCTAAGGCTGTGGGGTG	1260
QY	3015	gcaacttgatgatgatgatcatcctagtggttattacaggtgctcttlaaactgtaactgctc	3072
Db	1261	GCAATTTGATGATGATGATGATCTGAGTGTATTATTCAGAGTGCTTTTAAACTAGTAACTGCT	1318
RESULT	8		
AY027788	AY027788	1395 bp	MRNA
LOCUS	AY027788	1395 bp	MRNA
DEFINITION	Homo sapiens CLANB (CLANB1) MRNA, complete cds.		
ACCESSION	AY027788		
VERSION	AY027788.1	GI:14324114	
KEYWORDS			

SOURCE	ORGANISM	human.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 1395)	
AUTHORS	Damiano,D.S., Stehlik,C., Plo,F., Godzik,A. and Reed,J.C.	
JOURNAL	Clan, a novel human ced-4-like gene	
MEDLINE	Genomics. 75 (1-3), 77-83 (2001)	
PUBMED	21365712	
REFERENCE	11472070	
AUTHORS	2 (bases 1 to 1395)	
JOURNAL	Stehlik,C., Damiano,J.S., Plo,F., Godzik,A. and Reed,J.C.	
REFERENCE	Submitted (21-FEB-2001) Program on Apoptosis and Cell Death Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA	
FEATURES	Location/Qualifiers	
Source	1..1395	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="2"	
	/map="2p22-p21"	
	/lssuse_type="lung"	
gene	1..1395	
	/gene="CLAN1"	
	277..1356	
	/gene="CLAN1"	
	/codon_start=1	
	/product="CLANB"	
CDS	/protein_id="AAK14777.1"	
	/db_xref="GI:14324115"	
	/translation="MNFKDNRRALIORGMATVIRKQITDDLLFVNNVLRREVNITCEI	
	KVEDADARGIHMLIKKSGESCNFLSLKEMNPLPODNGSGLSGNLNLTK	
	LIMONIKMNEEDAIKLEGLKNLKMCLPFIHLISDGEKMDYVKSLSSEPOLDEI	
	OIVSCANNAVAKITLAONLHNVKSLTDSLSENYLEDKNGEALHELIRNNVLEQLTA	
	LMNPMGDCVNGSLSLKHLFEVQYIKLGLKNNRLDTEIRILGAFGKNNPLKNPO	
	LMNAGNVSSDGLTAFMGVFEENLKQLVFFDFSTKEFLPDALVTKLSQVLSKLFLOE	
	ARLVGWFFDDDDLSVITGARFLVTA"	
BASE COUNT	436 a 248 c 327 g 384 t	
ORIGIN		
Query Match	26.5% Score 815.4; DB 9; Length 1395;	
Best Local Similarity	95.9% Pred. No. 5.2e-208;	
Matches 837; Conservative	0; Mismatches 36; Indels 0; Gaps 0;	
QY	tcgtgaacaacccggaaaccttgatgatactgacacagaatcaacgagcgcgcgagt	2259
Db		
481	TTTCTTAATGCCCTTAAGAGAGTGAACATATCTCTATTTCAGACTTAATGACAAAGT	540
QY	ggctgactgacagcttgggtgaacttgaagaaccccttacaagctcotaatgaatgaata	2319
Db		
541	GGCTGACTGACAGCTTGCGTAACCTTGGAAGAACCTTCAAGAGCTCATATGATTAACATA	600
QY	aaagatgaatgaagaagatagctataaactagactgaagcctggaanaacccctgaagaagatg	2379
Db		
601	AAGATGAATGAAGAAGATGCTATTAAACTAGCTGAAGCCTGAAAACCTGAAGAGATG	660
QY	tgattattcatcttgaccacattgctcagacttggaagagggaatggaattacatagctcaag	2439
Db		
661	TGTTATATTTCATTGGACCCACCTTGCTGACATTGAGAGGGAATGATTACATAGTCAAG	720
QY	tctctgtcaaatggaaccccttggagacttgaagaataatcaatagatctcctgtgcttgcct	2499
Db		
721	TCTGTGTAAGGGAACCTGTGACCTTGTAAGAAATTCATTAGTCTCCTGTGCTGTGCT	780
QY	gcaaatgcagtgaaatcctctagctcagaactcttcaaatcttggtcaaacctgaacatcct	2559
Db		
781	GCAAAATGAGGGAATTCCTAGCTCAGATCTTTCACATTTTGTCGTAACATGACGATCTTT	840
QY	gattatcaagaataatctctgaggaagaagaatggaatgaagctcttcatgaaactgacgac	2619
Db		
841	GATTTATCAGAAAATATACCTGGAAGAAAGATGAATGAAGCTCTTCAATGAACTGATCGAC	900

QY	2620	aggatgaacgctctgaagaacgctcaacgcaactgacgctgacccctgggagctgtagcgtgcaag	2679
Db	901	AGGATGAACGCTCTGAAACAGCTCACCGCAGCTGATGCTGCCCTGGGCGCTGTACG	960
QY	2680	ggcagcctggagagcctctgtgaacacatttggagagaggtccccaactcgtcaagcttggg	2739
Db	961	GCGACCTGAGCAGCCTGTGTAAACATTTGGAGAGGTCCCAACATCTGTCAAGCTTGGG	1020
QY	2740	ttgaaaactggagagactcaacagatcagagatagaaatttagtgcatttlttgaaag	2799
Db	1021	TTGAAAACCTGGAGACTCACATATCAGAGATAGAAATTTTAAAGTGTGCAATTTTTC	1080
QY	2800	aacctctgaaaacattccagcagtttgaatttggcgggaaatcgtgtgagcagtg	2859
Db	1081	AACCTCTGAAAACCTTCCAGCAGTGTGAATTTGGCGGAAATTCGTGTAGCGAGTGT	1140
QY	2860	tgagctgcctcatggtgtgtattttagaattcttaagcaattagtgattttagacttagt	2919
Db	1141	TGAGCTGCCTCATGGGTGTATTTTGAATATTTAAGATCTTAAGCAATTTAGTGT	1200
QY	2920	actaagaattctctaccgcatccagcatatgctcagaaaacttagcgaagcttctc	2979
Db	1201	ACTAAGAATTCTCTACCTGATCTCACCATTTAGTCACAAAACCTTAGCCAAAGTGTAT	1260
QY	2980	ttaactttcttgcaagaagctagcgtcttgggtggcgaatttgatgagatgagatctcagt	3039
Db	1261	TTAAGCTTTCTGCAAGAACGTAGCGCTGTGGGTGGCAATTTGATGATGATGATCTCAGT	1320
QY	3040	gtattacaggtgctcttcaactagtaactgct	3072
Db	1321	GTTATTACAGGTGCTTTTAACTAGTAACTGCT	1353
RESULT	9		
LOCUS	AY027789	768 bp	20-JUL-2001
DEFINITION	Homo sapiens CLANC (CLAN1) mRNA, complete cds.		
ACCESSION	AY027789		
VERSION	AY027789.1	GI:14324116	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Hmo.		
AUTHORS	1 (bases 1 to 768)		
TITLE	Damiano, J.S., Stehlik, C., Pio, F., Godzik, A. and Reed, J.C.		
JOURNAL	Clan, a novel human ced-4-like gene		
MEDLINE	Genomics. 75 (1-3), 77-83 (2001)		
PubMed	21365712		
REFERENCE	11472070		
AUTHORS	2 (bases 1 to 768)		
TITLE	Stehlik, C., Damiano, J.S., Pio, F., Godzik, A. and Reed, J.C.		
JOURNAL	Direct Submission		
FEATURES	Submitted (21-FEB-2001) Program on Apoptosis and Cell Death Research, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA		
source	Location/Qualifiers		
	1..768		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="2"		
	/map="2p22-p21"		
	/tissue-type="lung"		
	1..768		
	/gene="CLAN1"		
	277..747		
	/gene="CLAN1"		
	/codon_start=1		
	/product="CLANC"		
	/protein_id="AAK14778.1"		
	/db_xref="gi:14324117"		
	/translation="MNEIKDSRALIORMGVTVIKQITDLEFVN/LNREVNICE"		

BASE COUNT	218 a	157 c	180 g	213 t	
ORIGIN					
Query Match	15.0%:	Score 461:	DB 9:	Length 768:	
Best Local Similarity	100.0%:	Pred. No. 9.7e-113:			
Matches 461:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:	
OY	1	atgaatttcataaagagacaatagtcgcgaagcccttattccaagaatggaatgactgtata	60		
DB	277	ATGAATTTCAATAAAGACACATATAGCCGAGCCCTATTATCAAGAATGGAGATGACTGTATA	336		
OY	61	aagcaatcacagatgacctattgtatgaaatgttctgaatcgcgaaagtaaacatc	120		
DB	337	AAGCAAAATCACAGATGACCTATTATTATGGAATGTTCTGGAATCGCGAAGAAATAAACATC	396		
OY	121	attctctcgagaaaggttgagcaggaatgcctgctaagagagatcatccaatgatlctgaa	180		
DB	397	ATTTCCTCGAGAAAGATGGAGCAGATCTGCTAAGAGGATCATATTCACATGATTTTGAA	456		
OY	181	aagggttcaagatccctgtaacctcttcttaatcccttaagaagtggaactatcccta	240		
DB	457	AAGGTTGAGATGCTCGTAACCTCTTCTTAATCCCTTAAGAGATGGAACATATCTCTTA	516		
OY	241	tttcagagacttaaatgagcaaaagctctttcttcacagacacacagagagacttgagcat	300		
DB	517	TTTCAAGACTTGAATGAGCAAAAGTCTTTTTCATCAGACATCAGAGGAGAGACTTGGACAT	576		
OY	301	ttagctcagagatttaaaagacttgcacataccacatcttctgaactttatccctt	360		
DB	577	TTGCGCTCAGAGATTTAAGAGACTTGTACCATACCCCATCTTTTCGAACTTTATGCCCTT	636		
OY	361	ggtgaagatattgacatatatttctaacttgaagaagaccctcacagaacctgctctg	420		
DB	637	GGTGAAGATATTGACATTATTTTAACTTGAAAGACACTTTCACAGAACTGTCTCTGTGG	696		
OY	421	aggaagagaccaaaccatcacgcggtgtagagcagctgaccc	461		
DB	697	AGGAAGAGCAACACACATCACCGCTGTGAGCAGCTGACCT	737		
RESULT 10					
LOCUS	AC011232/c				
DEFINITION	Homo sapiens chromosome 2 clone RP11-7BE13, WORKING DRAFT SEQUENCE.				
ACCESSION	AC011232				
VERSION	AC011232.7				
KEYWORDS	HTG; PHASE1; HTGS-DRAFT; HTGS-FULLTOP.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 185281)				
JOURNAL	Waterston, R.H.				
REFERENCE	The sequence of Homo sapiens clone				
AUTHORS	unpublished				
JOURNAL	2 (bases 1 to 185281)				
TITLE	Waterston, R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (04-OCT-1999)				
JOURNAL	Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
COMMENT	MO 63108, USA				
	On Mar 10, 2001 this sequence version replaced gi:9799811.				
	----- Genome Center -----				
	Center: Washington University Genome Sequencing Center				
	Center code: WUGSC				
	Web site: http://genome.wustl.edu/gsc/index.shtml				
	----- Project Information -----				
	Center project name: H.NH007BE13				

----- Summary Statistics -----

Sequencing vector: M13: 57%

Chemistry: Dye-terminator Big Dye; 50% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 182798 bases at least Q40

Consensus quality: 183538 bases at least Q30

Consensus quality: 184045 bases at least Q20

Insert size: 168000; agarose-fp

Insert size: 184681; sum-of-contigs

Quality coverage: 7.07 in Q20 bases; sum-of-contigs

Quality coverage: 7.33 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1985: contig of 1985 bp in length

1986 2085: gap of unknown length

2086 8183: contig of 6098 bp in length

8184 8283: gap of unknown length

8284 22741: contig of 14458 bp in length

22742 22841: gap of unknown length

22842 41446: contig of 18605 bp in length

41447 41546: gap of unknown length

41547 69182: contig of 27636 bp in length

69183 69282: gap of unknown length

69283 112131: contig of 42849 bp in length

112132 112231: gap of unknown length

112232 185281: contig of 73050 bp in length.

185281 Location/Qualifiers

FEATURES

source 1. 185281

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/clone="RP11-78E13"

1. 1985

/note="assembly_name:Contig11"

2086. 8183

/note="assembly_name:Contig12"

8284. 22741

/note="assembly_name:Contig13"

22842. 41446

/note="assembly_name:Contig14"

41547. 69182

/note="assembly_name:Contig15"

69283. 112131

/note="assembly_name:Contig16"

112232. 185281

/note="assembly_name:Contig17"

BASE COUNT 52719 a 36957 c 37187 g 57816 t 602 others

ORIGIN

Query Match 9.5%; Score 292.2; DB 2; Length 185281;

Best Local Similarity 85.9%; Pred. No. 5.5e-67;

Matches 324; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1972 gaattcgaactctgaggtcacactccggagattcgaagaattgaataagaagatc 2031

DB 185281 GAATTCGAGCACTCGAGGTCACACTCCGGGATTTCAGCAAGTTGAACCAAGATATC 185222

QY 2032 acatatcgggaaataatcagctcgcacaaagcctcaagctgcgaataaagaagatgt 2091

DB 185221 AGATATCTGGGGAAAATATTCAGCTTCGCCACAAGCCTCAAGCTGCATAATTAAGATGT 185162

QY 2092 gctgtgtgctggaaagcctcagttgtctcctcagcacctgttaagaacattatctctc 2151

|||||

DB 185161 GCTGTGTCGTGAGAGCCCTCAGTTTGTCCTCAGACACCTGTAGAACATTATTCCTC 185102

QY 2152 atgtgtgaagcagtcctccatcatagaatgagagacatcactcgttaacaac 2211

DB 185101 ATGTGGAAGCCAGTCCCTCACCATAGAGATGAGAGGACATCAGTCGTGAACAAAC 185042

QY 2212 ctgaaaccttgatcatcagacatcagaatcaagagcgcgggtggtcactaac 2271

DB 185041 CTGAAACCTTGATTCATCAGACCTACAGATCAACGCGTCCGGGATTGTGAATATTC 184982

QY 2272 agcttggtaactcgaagaccttcaaacgtcataatgataaataaagatgaatga 2331

DB 184981 AGTGTTGTGCTGTGTGCACTTAACAAAATAACAGTGTATAAATTGGAAAAAGAGA 184922

QY 2332 gaagatgctataaac 2348

DB 184921 GAGGAGACTTATTTCT 184905

RESULT 11

CNS01DS8/c CNS01DS8 162692 bp DNA PRI 02-MAR-2000

DEFINITION BAC sequence from the SPG4 candidate region at 2p21-zp22 BAC 563N04 of RPEC11 library from chromosome 2 of Homo sapiens (Human), complete sequence.

ACCESSION AL121658

VERSION HTG: HTGS-DRAFT; SPG4 genomic DNA interval.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 162692)

AUTHORS Hazan,J., Fonkheche,N., Mavel,D., Paternotte,C., Samson,D.,

Artiguenave,F., Davoine,C.S., Cruaud,C., Durf,A., Wincker,P.,

Brothier,P., Catolico,L., Barbe,V., Burgunder,J.M.,

Prud'Homme,J.F., Brice,A., Fontaine,B., Hellig,R. and

Weissenbach,J.

TITLE Spastin, a novel AAA protein, is altered in the most frequent form

JOURNAL of autosomal dominant spastic paraplegia

REFERENCE 2 (bases 1 to 162692)

AUTHORS Nat. Genet. (1999) In press

TITLE Direct Submission

JOURNAL Submitted (23-FEB-2000) to the EMBL/GenBank/DBJ databases

COMMENT On Mar 6, 2000 this sequence version replaced gi:6002391.

FEATURES Location/Qualifiers

source 1. 162692

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/clone_id="RPEC1-11"

/clone="563N04"

BASE COUNT 44981 a 32017 c 32887 g 52792 t 15 others

ORIGIN

Query Match 9.5%; Score 292; DB 9; Length 162692;

Best Local Similarity 100.0%; Pred. No. 6.1e-67;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2781 aggtgcatcttttggaaagaccctcgaanaactccagcagttgaatttgggggaa 2840

DB 152218 AGGTGCATTTTGGAAAGAACCTCTGAAAACCTCCAGCAAGTTGAATTTGGCGGAAA 152159

QY 2841 tcgtgtgagcagtgatgagtgctgcctcactgagtggttacttgaacttaagcaat 2900

DB 152158 TCGTGTCAGCAGTATGATGATGGCTTGCTCATGGGTATATTGAGATCTTAAGCAAT 152099

QY 2901 agtgttttcttacttactaagaattctcactgataccagcatatgcagaacct 2960

DB 152098 AGTGTTTTTACTTACTATAAGAAATTTCTACCTGATCAGCATAGTCAGAAAACT 152039

|||||

QY 2961 tagcgaagtgtatccaaagttaactttctcgaagaagcttaggcctgtgtggtggaatt 3020
 |||||||
 Db 152038 TAGCCAAAGCTTATCCAGCTTACTTTCTCGCAAGAGCTAGGCTGTGGGTTG:AAIT 151979
 QY 3021 tgaatataatctcagtggtatctacaggtgtctttaactagtaactgcct 3072
 |||||||
 Db 151978 TGATATGATGATCTGATGTTATTAACAGGTCCTTTTAACATAGTACTGCT 151927

RESULT 12
 AY027790 578 bp mRNA PRI 20-JUL-2001
 LOCUS Homo sapiens CLAND (CLAN1) complete cds.
 DEFINITION
 ACCESSION AY027790
 VERSION AY027790.1 GI:14324118
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 578)
 AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
 TITLE Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
 JOURNAL Genomics. 75 (1-3), 77-83 (2001)
 MEDLINE 21365712
 PUBMED 11472070

REFERENCE 2 (bases 1 to 578)
 AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
 Research The Burnham Institute, 10901 North Torrey Pines Road, La
 Jolla, CA 92037, USA

FEATURES
 source
 location/Qualifiers

1..578
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2p22-p21"
 /tissue_type="lung"
 1..578
 /gene="CLAN1"
 277..555
 /gene="CLAN1"
 /note="CARD protein"
 /codon_start=1
 /product="CLAND"
 /protein_id="AAK14779.1"
 /db_xref="GI:14324119"
 /translation="MNFIKDNRSLIORMGMTVIKQITDDLFVNVILNREEVNIICSE
 KVEDDARGRITHTIKKGSFSCNFKSLKEMNYPPLFQDLNGQSLVTA"

BASE COUNT 172 a 106 c 143 g 157 t
 ORIGIN

Query Match 8.8%; Score 269; DB 9; Length 578;
 Best Local Similarity 100.0%; Pred. No. 4e-61;
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaattcataaagaacatagccgagccctatccaagaatgggaatgactgtata 60
 |||||||
 Db 277 ATGAATTTCAATAAGGACATAGCCGAGCCCTTATTTCAAGAATGGGAATGATGTTATA 336
 QY 61 aagcaaatcacagatgacctattgtatggaaatgtctgaatcgcgaagaaglaaa 120
 |||||||
 Db 337 AAGCAAAATCACAGATGACCTATTGTATGGAATGTTCTGAATCGGAAGAAGTAA:ATC 396
 QY 121 attgtcgcgaagaagtgagcagatgctgtataaggagatcacaatgatttttaaa 180
 |||||||
 Db 397 ATTGCTCGAGAAAGTGAGAGATGCTGCTAGAGGATCATTCACATGATTTT:AAA 456
 QY 181 aaggttcagagtcctgtacacctcttctaatacccttaagaagatggaatactcc 240
 |||||||
 Db 457 AAGGTTCAAGATCCTGTAACCTTTTCTTAATCCCTTAAGAGAGTAATATCC:CTA 516

QY 241 ttccagacttgaaatgacaaagctttt 269
 |||||||
 Db 517 TTTCAGAGCTTGAAATGACAAAGCTTTT 545

RESULT 13
 G55568/c 553 bp DNA STS 30-MAR-2000
 LOCUS SHGC-100923 Human Homo sapiens STS genomic, sequence tagged site.
 DEFINITION
 ACCESSION G55568
 VERSION G55568.1 GI:6120887
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 553)
 AUTHORS Olivier,M. and Cox,D.R.
 TITLE Unpublished, Olivier, M., Cox, D.R. (2000)
 JOURNAL Unpublished (2000)

COMMENT
 Contact: Michael Olivier, David R. Cox
 Stanford Human Genome Center
 Stanford University School of Medicine
 4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
 Tel: (650) 320-5800
 Fax: (650) 320-5801
 Email: olivier@shgc.stanford.edu
 Primer A: AATTAAGGCGCAATAAATAGCAAA
 Primer B: TACACCCCTTTGTTCTTCCTCA
 STS size: 322

PCR Profile:
 Initial incubation: 95 degrees C for 10 minutes
 Denaturation: 94 degrees C for 30 seconds
 Annealing: 60 degrees C for 30 seconds
 Polymerization: 72 degrees C for 23 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9700

Protocol:
 Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 Amplifrag Gold Polymerase: 0.07 units/ul
 Total Vol: 5 ul

Buffer:
 MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3

FEATURES
 source
 location/Qualifiers

1..553
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="2"
 /clone_11b="Human"
 STS
 primer_bind 5..27
 primer_bind 5..326
 BASE COUNT 170 a 107 c 114 g 162 t
 ORIGIN

Query Match 5.5%; Score 170; DB 11; Length 553;
 Best Local Similarity 100.0%; Pred. No. 1.7e-34;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2615 tcacagatgaacgtgtcagacagctcacccagatgctgcctcgtggtgacg 2674
 |||||||


```

exon      complement(79055. .79174)
          /gene="Naip3"
          /note="region of sequence similarity; not known to be
          transcribed"
          /number=7
exon      complement(90835. .90887)
          /gene="Naip3"
          /note="region of sequence similarity; not known to be
          transcribed"
          /number=6
exon      complement(91714. .91794)
          /gene="Naip3"
          /note="region of sequence similarity; not known to be
          transcribed"
          /number=5
exon      complement(94865. .94966)
          /gene="Naip3"
          /note="region of sequence similarity; not known to be
          transcribed"
          /number=4
exon      complement(95293. .95983)
          /gene="Naip3"
          /note="region of sequence similarity; not known to be
          transcribed"
          /number=3
exon      complement(106079. .106141)
          /gene="Naip3"
          /note="region of sequence similarity; not known to be
          transcribed"
          /number=2
BASE COUNT 35695 a 24748 c 24777 g 32571 t
ORIGIN
Query Match 2.2%; Score 68.4; DB 10; Length 117791;
Best Local Similarity 53.0%; Pred. No. 7.8e-07;
Matches 197; Conservative 0; Mismatches 166; Indels 9; Gaps 2;
OY 500 ttgaagggaatctgcaaaagcaagtcacactctgctgcagcagcatgcatgctcggg 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17517 TGGAGGGGGAAGCTGGAAGGGAAGCAACCTTCTGAGAGAGATGCTTTCTCTGGG 17458
OY 560 gctcgggaagctgcaagcgtctgcaagtcacactctgctgcagcagcatgcatgctcggg 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17457 CATCAGAGATGCTGCCCCCTGTTGTACAGGTTCCACATGCTTCTTACTCTCTCC CTAGT 17398
OY 615 -cagcagggcccgaggtgagcttttgaacccctctgtagcaactctgagatatactcg 673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17397 CCATCACACCAAGTACAGGAGCTGCGCAACATCTGTGTACCAACTCTTACGCG:CAAGGAG 17338
OY 674 gcaacatcaggaagcaagcaacttcaagccatgctgctggaagctggcggaaggtctctt 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17337 GCGGCACTTACTGAGTGTGCTGACGACAGCATCCAGCATTTACAAACACAGTGTCTGT 17278
OY 734 tctctcttgatgctacaatgaatltcaagcccaagaacgtgccgaagaatcgaaag vccctga 793
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17277 TCGCTGTGGATGCTACAGAGGGCT--GGCCCTCACTCCCAAGCCCTACACA(AC)TGA 17221
OY 794 taaagaaaacaacagcgttcaagaacatgctacgtcaccaactgaagtgactga 833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17220 TTACAAAACAACTACTGTTCCGAGACCTGCTTATTGATTCATACAAACAG:gtTCA 17161
OY 854 ggcacatacggc 865
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17160 GAGACATCCGCC 17149

```

```

VERSION AF135490.1 GI:5932007
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4752)
AUTHORS Huang,S., Scharf,J.M., Growney,J.D., Endlitzl,M.G. and
          Dietrich,W.F.
          The mouse Naip gene cluster on Chromosome 13 encodes several
          distinct functional transcripts
          Mamm. Genome 10 (10), 1032-1035 (1999)
TITLE 2 (bases 1 to 4752)
JOURNAL Dietrich,W.F. and Huang,S.
MEDLINE 99431676
REFERENCE 2 (bases 1 to 4752)
AUTHORS Dietrich,W.F. and Huang,S.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1999) Genetics, HHMI/HMS, 200 Longwood Ave,
          Boston, MA 02115, USA
FEATURES
source location/Qualifiers
1..4752
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="21"
/cell_type="primary macrophage"
1..4752
/gene="Naip"
113..4456
/gene="Naip"
/note="Naip-rs6"
/codon_start=1
/product="neuronal apoptosis inhibitory protein"
/protein_id="A056762.1"
/db_xref="GI:5932008"
/translation="MAOGEAVEEIIICEFDDLVSELTSLRVDAALSVLKROOEDHK
          /TNMKMGKGNFSOMSEARLKTETFEFLGKPRSTMPQPMAGAFHTGKVGCCFCSL
          ILESTEDLPFAHGTSPRLVSAAGVFGKRDYVOCFSGGCLGMEBGGDDPMKHA
          KMPKCEFLQSKSPREIRQYVSYGFLVGHENHVMVREELPVAVACNDSVFA
          NEELRMDTFKDMHPESPAVEALVKAGLRYTGKRDYVOCFSGGCEKMAKEDQPIED
          HTRFPCNVFLQTKSAVITPALQSHCALPEMETETSSNDDAAVMSIVADVDS
          EAOELPASPASIVSLCRDODHSPAGRCAGSGATYPTSDLGSEDMOAEKRSSEO
          LBDYTKATFRHNMNLPPEVYSLGTDHLLCDVSIISKHISOPVQGLTIPVEVSNLS
          VMCVEGASGKTTFLKRIAPLWASGCCULNRFOPVFLSSTIPGOELAKITCAO
          LGAGGCISEVCSSTIIOLOHOVFLVLDYSGLASLPALHTLITKLSRCLLIA
          VHNKVRGIRPYLDTSLTEIKPEFVTVVLRKLFSDHIMYAKFLNYGFSBELGIG
          HRTPLFVAAVCTDMFNKPSDOPDVALRKAWOYSLRKHKAKRFLQATVSSCOIA
          LRGIFSSCFEPNSDNLAEAGVDEDELTCIMSKFTFAORLPYRFLRGLDFELAAV
          RLTELLSDROEDODGLYIROIINSLPKAMSTYHTELFYVSSHPSSKAPVYSHL
          OLVDKESIEENNEENDYMKIHDEALMLTECRGIMOLSPEPSLFTISENLRICTNF
          AHESNTVACSPVITLOFLRGRTLDLKVLSLOYPMHDPRLTLLIKSTIKSLNNMVR
          IDESLIEKPEVQOPTIDODVALIAROPINENOKNSKKHITIKYEDMKHOLPNTS
          TGVWKSIPRPKTKPKLEUYVNTNGADOLLOVLMVPSASGISEPRLSDGSELEST
          FALVELSKASTKCSNRISREDOKLITLOSTLOSIVASGNOIPDLPHNHLFEI
          GLAEIDVRDLSKRDVLSVYIGERFPHNHEKISITRSTSDSKIKLLONSNHLHF
          HUKCNFLSKCEPLATYVADGKRLRELEISGRCGEATFTVAILNPEFLVILNRDQF
          POKETSEKTRQALGSLRNLEKLVYPGIDGHOVAKLIVQCIDLPCLRVLVREITDD
          DSVLEIAKATRGCFORLENDDTLNHTTTERGYRNFVYLDMLPKMNDLSRIHPE
          CIOIOATVYKALGQCVSRILPSITRLGLMSLIDDEEDIKVINDYKERRHPQSKRLTVHWR
          WVPESEPVIOK"
BASE COUNT 1313 a 1111 c 1114 g 1214 t
ORIGIN
Query Match 2.2%; Score 67.8; DB 10; Length 4752;
Best Local Similarity 56.7%; Pred. No. 7e-07;
Matches 148; Conservative 0; Mismatches 107; Indels 6; Gaps 1;
OY 500 ttgaagggaatctgcaaaagcaagtcacactctgctgcagcagcatgcatgctcggg 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1647 TGGAGGGGGAAGCTGGAAGGGAAGCAACCTTCTGAGAGAGATGCTTTCTCTGGG 1706
OY 560 gctcgggaagctgcaagcgtctgcaagtcacactctgctgcagcagcatgcatgctcggg 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 1707 CATCAGATGCTGCCCGCTGTGAACAGGTTCCAGCTGCTTCTTACCTTCCTTAGTT 1766
Oy 614 tcaagaggccccaggtgacttttgaaccctctgtatcaactcctlgatatacctg 673
Db 1767 CCATCACACCAAGCCAGAACTGGCCAAATATCTGTGCCAAGCTCCTAGGGCAGAG 1826
Oy 674 gcacaatcaggaagcagacatcattcatgccaatgctgaagctgcgcaggaaggtcctt 733
Db 1827 GCTGCATTAGTGAAGTGTGTCTGAGCAGCATTCACAGAGTTACACACCAAGTGTGT 1886
Oy 734 tccttcttgatggctacaatg 754
Db 1887 TCCTGTGATGACTACAGTG 1907

Search completed: March 25, 2002, 13:24:16
Time: 9240 sec

THIS PAGE BLANK (USPTO)


```

QY 323 LQIQSRCLRNLMKTPLEFVITCAIOMGESEFHSHTOTLFTFTFDLLLOKNNHKKHGYA 382
Db 627 VYFGKQSLQKTKTLPFAALCAHMFQYPPDPSPDDVAFKSYMERLSLRN-----A 680
QY 383 ASDFLR-SLDHGGDLAEGVFSHKDFELOY--SSVNDVLLTGLCKYTAJRKPKY 439
Db 681 TAILATVSSCGELALKFSCCFEFDNDDLAEGVDEDEDLTKLMSKFTAJRLPEY 740
QY 440 KPEHKSFOETAGRLSLTSHPEEVTGNGYLOKMWISIDITSTYSLLRTGGSV 499
Db 741 RFLSPAFQELGMRILIELSDROHODGLYHLKQINSPMTVSAYNNFLNV--SSL 798
QY 500 EATRA---VMKHLAAYOHCGLGISIAKRLPMLROESLOSVKN-----TTEOET 545
Db 799 PSTKAGPKIVSHLLHVDN-----KESLEINSEDDYLKHQPE:SLQMO: 843
QY 546 LKAI-----NINSEFECGILH-----VOESTSKSALSOEFAPFGKSLYITSGNIPD 593
Db 844 LRGLMQLCPQAVFSVSEHLVLAALKTAVOGNT--VAACSPFVLQFLOGRTLLGLALNT-Q 901
QY 594 YLPD-----FPEHLPLNC--ASALDFIKLDFYGG--AMASWE 625
Db 902 YFPDHEESLSLRSHFSIRGNTSPRAHFSVLETCFDSQVPTIDODVYASAFEPNEME 961
QY 626 -----KAEPTGGINHMEAPETY-IPSRVSLFPFNKQEFRTLEYTL 666
Db 962 RNLAEKEDNKSVMQORRASPDLSTGYWKLSPKQYKIP-----JLEVDY 1006
QY 667 RQFSKLKODITYLGIKIFSSATSLRQIKRCAGVAGSLVSTCK-NIYSLMV:ASPLT 725
Db 1007 NDDIVYGQDMLELMTVFSASQRIELHNSRGFTESIRPALELSKASVTKSIS:KLELS 1066
QY 726 IEDERHITSVTNKTLSIH---DLQNRPLPGCLTDSLGNKNTKILMINKM----- 775
Db 1067 AAEQELLTLPLSLEVSCTIOSODIPEFN--LDKFLCKELSLVDEGINVF:VIPEE 1124
QY 776 -----NEEDAIIKLAEGIKNLKKKCLFHLTHL---SDIGECMDYIVKSL 816
Db 1125 FPNFHHMEKLLQISAEYPSKLVASLPNFSILKILNLEGOQFPDEETSEKAVYJLGSL 1184
QY 817 SEPCDLEELQVSCCLANAVKILAOHLNLVYKLSIL 853
Db 1185 ---NLEELLPTGDCIYVAKLIIQCCOCLHCLRLV 1217

RESULT 2
US-08-836-134-23
Sequence 23, Application US/08836134A
Patent No. 6020127
GENERAL INFORMATION:
APPLICANT: Mackenzie, Alex E.
APPLICANT: Korneluk, Robert G.
APPLICANT: Mahadevan, Mani S.
APPLICANT: McLean, Michael
APPLICANT: Roy, Natalie
APPLICANT: Ireda, John
TITLE OF INVENTION: Neuronal Apoptosis Inhibitor Protein, Gene Sequence and
Patent No. 6020127
TITLE OF INVENTION: Mutations Causative of Spinal Muscular Atrophy
FILE REFERENCE: 3477-112, 033477/139914
CURRENT APPLICATION NUMBER: US/08/836,134A
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 23
LENGTH: 1151
TYPE: PRT
ORGANISM: Homo sapiens
US-08-836-134-23

```

Query Match 10.2%; Score 540.5; DB 3; Length 1151;
Best Local Similarity 24.5%; Pred. No. 3.1e-40;

```

Matches 204; Conservative 144; Mismatches 300; Indels 185; Gaps 30;
QY 69 ELKSLKEMNYPLFQDINGO-----SLPHQTSBGDLDD----- 100
Db 353 FQNNKMS--SAEYTPDQSGEJCCELETTSESNEISAVGPIVBEAAGQEMFOEAKN 411
QY 101 LAQDLKDLVHTPSFLNF-----YPLGEDIDITFNKSTFTTEPYLMRKDOHNHR 148
Db 412 LNEQLAAVTSASFRRMSLLDISDLPATDHLGCDIST--ASKNISKV----- 458
QY 149 VEDLTNGLLQALQSPCIIIEGSGKSTLLQRIAMLMGSGCKALTKREVFYLRLS-- 206
Db 459 QEPLVLPVEFGNLSVWCVEGAGSGKTVLKRKIFLMAWGCCPLNRPOLVFYLSLST 518
QY 207 RAQGLFETLCDOLDIPGTIRKQTFEMAMLLKRLORVFLDLGYNDF--KPNCEIEAL 264
Db 519 RPDEGLASTICDQLLEKESGVTEMCMRNITIQDLKNQVFLFLDYEKICSIQ--VIGRL 575
QY 265 IKENHFRKMVIVTTTECLRHIFRGALTAFCWMTEDSAQALIREVILKELA--EGLL 322
Db 576 LQKNSLRCLIAVNTNRADIRRYLETILQAFEPFYNYCIRKLFSHMNTLRKFM 635
QY 323 LQIQSRCLRNLMKTPLEFVITCAIOMGESEFHSHTOTLFTFTFDLLLOKNNHKKHGYA 382
Db 636 VYFGKQSLQKTKTLPFAALCAHMFQYPPDPSPDDVAFKSYMERLSLRN-----A 689
QY 383 ASDFLR-SLDHGGDLAEGVFSHKDFELOY--SSVNDVLLTGLCKYTAJRKPKY 439
Db 690 TAILATVSSCGELALKFSCCFEFDNDDLAEGVDEDEDLTKLMSKFTAJRLPEY 749
QY 440 KPEHKSFOETAGRLSLTSHPEEVTGNGYLOKMWISIDITSTYSLLRTGGSV 499
Db 750 RFLSPAFQELGMRILIELSDROHODGLYHLKQINSPMTVSAYNNFLNV--SSL 807
QY 500 EATRA---VMKHLAAYOHCGLGISIAKRLPMLROESLOSVKN-----TTEOET 545
Db 808 PSTKAGPKIVSHLLHVDN-----KESLEINSEDDYLKHQPE:SLQMO: 852
QY 546 LKAI-----NINSEFECGILH-----VOESTSKSALSOEFAPFGKSLYITSGNIPD 593
Db 853 LRGLMQLCPQAVFSVSEHLVLAALKTAVOGNT--VAACSPFVLQFLOGRTLLGLALNT-Q 910
QY 594 YLPD-----FPEHLPLNC--ASALDFIKLDFYGG--AMASWE 625
Db 911 YFPDHEESLSLRSHFSIRGNTSPRAHFSVLETCFDSQVPTIDODVYASAFEPNEME 970
QY 626 -----KAEPTGGINHMEAPETY-IPSRVSLFPFNKQEFRTLEYTL 666
Db 971 RNLAEKEDNKSVMQORRASPDLSTGYWKLSPKQYKIP-----CLEVDY 1015
QY 667 RQFSKLKODITYLGIKIFSSATSLRQIKRCAGVAGSLVSTCK-NIYSLMV:ASPLT 725
Db 1016 NDDIVYGQDMLELMTVFSASQRIELHNSRGFTESIRPALELSKASVTKSIS:KLELS 1075
QY 726 IEDERHITSVTNKTLSIH---DLQNRPLPGCLTDSLGNKNTKILMINKM----- 775
Db 1076 AAEQELLTLPLSLEVSCTIOSODIPEFN--LDKFLCKELSLVDEGINVF 1126

RESULT 3
US-09-180-439-3
Sequence 3, Application US/09180439
Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
APPLICANT: Hatzixanthis, Kostas
APPLICANT: Jones, David A
APPLICANT: Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180,439
EARLIER APPLICATION NUMBER: PCT/GB97/01249

```

us-09-697-089-2.rat

Tue Mar 26 11:36:23 2002

EARLIER FILING DATE: 1997-05-08
 EARLIER APPLICATION NUMBER: GB 9609681.3
 EARLIER FILING DATE: 1996-05-09
 EARLIER APPLICATION NUMBER: GB 9619924.5
 EARLIER FILING DATE: 1996-09-24
 EARLIER FILING DATE: 1996-05-09
 EARLIER FILING DATE: 1996-09-24
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO: 3
 LENGTH: 968
 TYPE: PRT
 ORGANISM: Lycopersicon esculentum
 US-09-180-439-3

Query Match 3.3%, Score 172, DB 4, Length 968, Gaps 38,
 Best Local Similarity 23.0%, Pred. No. 1.2e-06, Indels 176, Mismatches 208
 Matches 147; Conservative 108; Mismatches 208
 459 LSHPEEYTKNGYLOKMSIS-DITSYSSILRYTCSSVEATRAVKKHLAVYOHGC 517
 203 LSGFPEI-----GYLRSLTKLSDIN-----FLSGIPASIGNLNLSFLLYNN 249
 518 LGLSIAKRPMLROESLOSQVKNTEBOELKAININSFVECGIHLXOESTSKSA-----IS 572
 250 QLSGSI-----PELIGYLRSLTK-----LSLGINFLSGSIPASIGNLNLS 290
 573 OEFEAFQOKSLYIN--SGNIPD--YL-----PDFEER-----LPNCASALDPT-KLDF 616
 291 R-----LDLNNKLSGSIPEIGYLRSLTYLDLGENALNGSIPASIGNLNLSRDL 342
 617 YGAMASMEKAAEDTCGIM-----EAPETIYPSRAVSLFENKQEPRTLEYTLDF 669
 343 YNNKLSG--SIPETIGYLRSLTYLDLGENALNGSIPASIGNLNLSRDL 342
 670 SKLNKODIYLGKIFSSNTSLRQIKRCAGVAGSLVSTCKNIYSLWVSPLT--I 726
 396 SGIPEIETYL-----SSLTELYLGN--SLNGSIPASIGNLNLSFLLYNNOLSGSIP 448
 727 EDERHTSVNKLKTLSDINORLPGLDLSIGNIKNTKLMINDIKNNEDEAIKLAG 786
 449 EEIGYLSITEL-----FLGNNSLNGSIPASIGNLNLSRILYLYNNOLS--GSIPASFG 500
 787 -LKNKKMCLFHLTHSD-----IGEGMDYIVKSLSE-----PC-DLEEIQ 826
 501 NMRNLQTL-----FLSDNDLIGEIFSVCNLTSLVLYMSRNKLKGVPOCLGNISDLH 554
 827 LVSCCLSANVYK--ILAOINLHLYKSLIDLSSENYLEKD-----GNEALHELIDRNNVLEQ 880
 555 ILS--WSNSFRGELPSSISNLSIKILDFGRNNLEGAIPDFGNISLOVFDMON--NK 610
 881 LVALMLPMGCDVOCSSLSLKH-----LEEV-----OLVYLGKNNRLTDT----- 922
 611 LSG-TLPTNFSIGSLISLHNGELADEIPRSIDNCKKLOVLDLGDNDLNTDFPMMLGT 669
 923 --EIRILGAFPGK--NPLK-----NFQOLNLAGNRSVSDGMLAFMGVFNENKOLYF 969
 670 LPELRVLRITSNKLGPIRSSGAEIAPDRLITLDSRNAFSD--LPTSLFHLKGKMT 726
 970 FDFSTYKE-----FLDPPALV-----KLSOVLSKLTFL 997
 727 VDKTMEPSYESIYDDSVVVYTKGLELYRILSLYITII 765

RESULT 4
 US-09-180-439-4
 Application US/09180439

Sequence 4, Application US/09180439
 Patent No. 6225532
 GENERAL INFORMATION:
 APPLICANT: Dixon, Mark S
 APPLICANT: Halzixanthus, A
 APPLICANT: Jones, Jonathan DG
 APPLICANT: Jones, Jonathan DG
 TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

FILE REFERENCE: 620 - 53
 CURRENT APPLICATION NUMBER: US/09/180.439
 CURRENT FILING DATE: 1998-12-06
 CURRENT APPLICATION NUMBER: PCT/GB97/01249
 EARLIER FILING DATE: 1997-05-08
 EARLIER APPLICATION NUMBER: GB 9609681.3
 EARLIER FILING DATE: 1996-05-09
 EARLIER APPLICATION NUMBER: GB 9619924.5
 EARLIER FILING DATE: 1996-09-24
 EARLIER FILING DATE: 1996-05-09
 EARLIER FILING DATE: 1996-09-24
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: Patentln Ver. 2.0
 SEQ ID NO: 4
 LENGTH: 968
 TYPE: PRT
 ORGANISM: Lycopersicon esculentum
 US-09-180-439-4

Query Match 3.3%, Score 172, DB 4, Length 968, Gaps 38,
 Best Local Similarity 23.0%, Pred. No. 1.2e-06, Indels 176, Mismatches 208
 Matches 147; Conservative 108; Mismatches 208
 459 LSHPEEYTKNGYLOKMSIS-DITSYSSILRYTCSSVEATRAVKKHLAVYOHGC 517
 203 LSGFPEI-----GYLRSLTKLSDIN-----FLSGIPASIGNLNLSFLLYNN 249
 518 LGLSIAKRPMLROESLOSQVKNTEBOELKAININSFVECGIHLXOESTSKSA-----IS 572
 250 QLSGSI-----PELIGYLRSLTK-----LSLGINFLSGSIPASIGNLNLS 290
 573 OEFEAFQOKSLYIN--SGNIPD--YL-----PDFEER-----LPNCASALDPT-KLDF 616
 291 R-----LDLNNKLSGSIPEIGYLRSLTYLDLGENALNGSIPASIGNLNLSRDL 342
 617 YGAMASMEKAAEDTCGIM-----EAPETIYPSRAVSLFENKQEPRTLEYTLDF 669
 343 YNNKLSG--SIPETIGYLRSLTYLDLGENALNGSIPASIGNLNLSRDL 342
 670 SKLNKODIYLGKIFSSNTSLRQIKRCAGVAGSLVSTCKNIYSLWVSPLT--I 726
 396 SGIPEIETYL-----SSLTELYLGN--SLNGSIPASIGNLNLSFLLYNNOLSGSIP 448
 727 EDERHTSVNKLKTLSDINORLPGLDLSIGNIKNTKLMINDIKNNEDEAIKLAG 786
 449 EEIGYLSITEL-----FLGNNSLNGSIPASIGNLNLSRILYLYNNOLS--GSIPASFG 500
 787 -LKNKKMCLFHLTHSD-----IGEGMDYIVKSLSE-----PC-DLEEIQ 826
 501 NMRNLQTL-----FLSDNDLIGEIFSVCNLTSLVLYMSRNKLKGVPOCLGNISDLH 554
 827 LVSCCLSANVYK--ILAOINLHLYKSLIDLSSENYLEKD-----GNEALHELIDRNNVLEQ 880
 555 ILS--WSNSFRGELPSSISNLSIKILDFGRNNLEGAIPDFGNISLOVFDMON--NK 610
 881 LVALMLPMGCDVOCSSLSLKH-----LEEV-----OLVYLGKNNRLTDT----- 922
 611 LSG-TLPTNFSIGSLISLHNGELADEIPRSIDNCKKLOVLDLGDNDLNTDFPMMLGT 669
 923 --EIRILGAFPGK--NPLK-----NFQOLNLAGNRSVSDGMLAFMGVFNENKOLYF 969
 670 LPELRVLRITSNKLGPIRSSGAEIAPDRLITLDSRNAFSD--LPTSLFHLKGKMT 726
 970 FDFSTYKE-----FLDPPALV-----KLSOVLSKLTFL 997
 727 VDKTMEPSYESIYDDSVVVYTKGLELYRILSLYITII 765

RESULT 5
 US-08-930-996A-2
 Application US/08930996A

Sequence 2, Application US/08930996A
 Patent No. 6100449
 GENERAL INFORMATION:
 APPLICANT: Fluhr, Robert

APPLICANT: ESHED, Yuval
APPLICANT: ORI, Naomi
APPLICANT: PAPAN, Ilan
APPLICANT: ZAMIR, Daniel
TITLE OF INVENTION: A GENE FAMILY FROM THE 12 FUSARIUM RESISTANCE
TITLE OF INVENTION: LOCUS OF TOMATO AND USE THEREOF FOR TRANSFORMATION AND
TITLE OF INVENTION: SELECTIVE BREEDING OF TOMATO AND RELATED PLANTS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/930,996A
FILING DATE: 09-DEC-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05272
FILING DATE: 15-APR-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: IL 113,373
FILING DATE: 13-APR-1995
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-930-996A-2

Query Match 3.2%; Score 170.5; DB 3; Length 1220;
Best Local Similarity 19.8%; Pred No 2,4e-06;
Matches 231; Conservative 160; Mismatches 393; Indels 383; Gaps 61;

QY 9 RALIQKMGIVYQITDDLFVWVNLNREEVNIICCEVEDDARAGIIMLLKSGESC- 67
DB 242 KGLIQEIGSDLK-ADD-----NLNOLVKKLADNMLN-----LAQ LKDIYHT 111
QY 68 -LEKSLKE-W-NYPLFODLNGSIFHOTSEGLDD-----LQV LKDKLNG 285
DB 286 KRLPLVLDVWMDVNPWDL--RNLFLO--GDIGSKITVTRKESVALMDPSGATY-- 338
QY 112 PSLNLFYPLG---EDIDIIFNLSKSTFTEPVLMRKD-QHHHRYEQI-----T-NGI--- 157
DB 339 -----MILSSEDEWALFKRRHS-----LEHKDKEPEDEEVGKQIADK KGLPLA 384
QY 158 LQALQSPCIIIEGSGKSTLLQRIAMLWSGKCALTKRKFVFFLRLSBAQGLFETLC 217
DB 385 LKALAGLRKSEVDENRILRSEI--WELPSC-----SNGIIPALM 424
QY 218 DQLDIDIGTIRKQTFMAMLLKRLRQVLFLLDGYNEFRPQNCPELEALIKENHR- KMMVT 276
DB 425 LSYNDLPAHLKO-----CLAVCAIYPRDYQFRKDOVI 456
QY 277 VTTTECLRIIROFGA-----LFAVGDMTEDSAQALIREVLLIKELAGLLAIQKS 328
DB 457 HLMIANGLVH--OFHSGNOYFIELRSRSLFEMASEPSEBVEEFLMDLVNDLQIASSN 514
QY 329 RCLNKLKTPLEFVYITCAI-----QMGESB-----FHSHTQTL-----FHTFY LLI 371
DB 515 HCLRLDENKSGHMLQCRHMSYSIGDGEFEKRLSLFKSGLRTLLPLDIOFH-YSLKLS 573
QY 372 OKNHHKRGVAAPFISLHOC-----GDALGEGVSHKDFELQVSSVNEQVL----- 421
DB 574 KRVHNLPLTFLRSRLALSLSHYQIEVLPNDLFIRKLKRLFLDLSETSIITKLPSIFV.XN 633

QY 422 LTTGLL--CKYTAQKFKPKYKFFHKSFQETTAGRLS--SILTSHEPEVYKNGYL--- 474
DB 634 LETLLSCEVLEEDPLOMEKILINRHLIDISNTRKRLPLHLRSKQLVAGAKFLVG 693
QY 475 OKRVISIDTTSYSLRYTCSSV---EATPAVKHILAAVYOHCLLGLSIAPKPLRQ 531
DB 694 WRREYLGEPNLYGSLSTILELNVDRRAVAKMKENKHWQ-----LSLE---W-S 742
QY 532 ESLSQVKNTEOEIL-----KAININSVEGCIHLYOESTSKSALSQEFAPFOGK 582
DB 743 ESTSADNSQTERDILDELPRHKNIRAVETGY----- 774
QY 583 SLYNSGNIPLYFD-----FEHLPNCAS-----ALDFIKL----- 614
DB 775 ---RGTNFPNWVADPLFVKLVHLRLNCKDCYSLPALGQLPCLFELSRGHGIRVTE 830
QY 615 DRYGAMASMEKAEDTGIHMEAPETYIPRAVSLFENMKQ-----ERRLEVTLR 667
DB 831 ERYG--RLSKKRPNSVLVRLFEDMPE-----WQWHTLIGIEPTE--- 871
QY 668 DFKLKNQDITVGLKIFSSATSRLQIKRACVAGSLSLVST-----CKNIYSIMV 719
DB 872 -----KLSIKNCPELSLEIPFOFSLRLDIDCKSVTSFPF 908
QY 720 EASPLTE-----DERHITSVNTKTSIHDLQNRQLPGGLTDSLGNLK 763
DB 909 SLPTTLKRIKISCPKLEAPVGMFVEYLSVDCGCVDDISPEFLPTARQLSIEMCH 968
QY 764 NTKLIMDNIKNMEEDAIKLAGIKRLKMKCL-----FHLTHLSDIGEMDIYVKSLS 818
DB 969 NTRFLIPTATESLH-----IRNCKLSMACGGAQDTSLNIWG-----CKKLKCL 1014
QY 819 P--CGLFEIOIYSCCLSNVAVKILIAONLHNLVKSITLDSSENYLER--DGNEALHELID 873
DB 1015 PELLPLKELRLTYC-----PEIBELPFN--LOIDI--RYCKKLVNGRKEH--- 1059
QY 874 RMNVLQOLALMPW--GCVQGSLSLKLHLE--EYPOLVYKLGKRWRLTDEIRIGAF 930
DB 1060 ---LDRLEMLKTHGSD-----EHLEWELPSSIQ-----RLFTINIKLTSQ 1100
QY 931 FGNPNLKNQOLNLA--RVSSDGLAWGVFENIKOLVDFDFSKTEPLPPLVYKLS 988
DB 1101 HLKS-LTSLQFLRTVGNLSQFQSGQLSFSFSLTSLQTLQIMFNLQSLPSALPSSLS 1159
QY 989 Q-VLSKRLFLQEARLVGQFDDDLASY 1014
DB 1160 HLISNCPNLSQSLPLKGMPSLSLSTLSI 1186

RESULT 6
US-09-353-585-2
Sequence 2, Application US/09353585
Patent No. 6287865
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
Jones, David A
Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses
thereof
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6287865th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997
APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 9506658.5
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-353-585-2

Query Match 3.1%; Score 165; DB 4; Length 1112;
Best Local Similarity 21.6%; Pred. No. 6.6e-06;
Matches 139; Conservative 81; Mismatches 199; Indels 224; Gaps 31;

QY 425 GLICKYTAQRKPKYKFFHRSFOETAGRRSLSTSHPEEVTKGNGYLQKMSISDIT 484
DB 140 GLALAQIR-----IFHNO-----LNGFIPKEI---GYLSRLKLS-LG 175
QY 485 STYSILRYTCGSSEYATRAVMKHLAAVYOHGCLGLSIAKRPPLMROESLOSVAKNTTBOE 544
DB 176 INFPLS-----GSIPASVGNLNNISFLYLNNQLSGSI-----PEISYLSLTFBLD 221
QY 545 ILKAININSEVCEGITHLYOESTSKALSQEFDA-----FPOGKSLYINSGNIPD 593
DB 222 L-----SDNALNGSIPASISGNMNNLSFLYGNQL---SGSIPE 257
QY 594 ---YL-----FDPEHLPN---CASALDFIKLDF---YGGAMASWEKAEDTGGIHMEEA 639
DB 258 EICYURSLTYIDLSENALNGSIPASISGNLNNLSFLYGNQLSG--SIPEIG----- 308
QY 640 PETYIPSRVSLFPMKOEFTLEVTLLRDFSKLNKODITYLGKIFSSATSLRLQIKRCAG 699
DB 309 ---YL-----RSLNVLGSENALN----- 324
QY 700 VAGSISLVLSTCKNITYSLMVEASPLTJEDERHITSVTNLKTLSTHDLONQRLPGGLTDSL 759
DB 325 --GSIPASISGNLNNISFLYLNNQLSGSI-----PEISYLSLTFBLD 221
QY 760 GNLKNTLTLINDNFKMNEDEAKLAEGKLNKKMCLFHLTLHSDIGCEMDYIVR-----SL 815
DB 380 GNLNLSMLLYLNNQLSGSIPLASLG--NLNNLSRLYLNNQLSGSIPEIGYLSLTYIDL 438
QY 816 SSEPCD-----LEEIQVSCCLSANAVKILAOHLNVLVLSITLSDSEN 858
DB 439 SNNISNGIFPASFGNMSLAFLEYENQLASS-----VPEIGYLSLTLVNDLSEN 489
QY 859 YLEADGNALHELIDRMVLEQULRALMPWCCDVQGSISLILKHEEYPOVLVKGGLKMR 918
DB 490 AL---NGSIPASISGNLNNISRLNL-----VNNQLSGSIPEIGYLSLTLVNDL- 533

QY 919 LTPREIR-ILGAFPGKNPLKMFQOLNLAGNRVSSD-----GWL----- 955
DB 534 LSENALNGSIPASFGN--LNNLSRLNLVNNQLSGSIPEIGYLSLTLVNDLGLSENALNGSI 591
QY 956 -AFMGVPENLKQLVFDFSTKEFLPDPAIVRKRLSQVLSKTFEL 997
DB 592 PASISGNLNNLSMLYLNNQLSGSIPEI-----GYLSLTYL 628

RESULT 7
US-09-353-585-3
Sequence 3, Application US/09353585
Patent No. 6287865
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
Jones, David A
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th floor, 1100 No. 6287865th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997
APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 9506658.5
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-353-585-3

Query Match 3.1%; Score 165; DB 4; Length 1112;
Best Local Similarity 21.6%; Pred. No. 6.6e-06;
Matches 139; Conservative 81; Mismatches 199; Indels 224; Gaps 31;
QY 425 GLICKYTAQRKPKYKFFHRSFOETAGRRSLSTSHPEEVTKGNGYLQKMSISDIT 484
DB 140 GLALAQIR-----IFHNO-----LNGFIPKEI---GYLSRLKLS-LG 175

Tue Mar 26 11:36:23 2002

us-09-697-089-2.ra1

Page 6

Db 140 GILAKLQIR-----IFHNQ-----LNGFIPKEI-----GYLRSVTKLS-LG 175
QY 485 STYSSLLRYTCGSSVEATRAVMKHLAAVYOHGCLLGLSTAKRPMQESLOS/KMTTBOE 544
Db 176 INFIS-----GSPASVGNLNNLSFLYLYNNOLSGSI-----PEELISY:KSLTELD 221
QY 545 ILKAININSVEEGCIHLHYOESTKSALISOEFEA-----PFOCKSLY:MSGNIPD 593
Db 222 L-----SDNALNGSIPASIGNMMNNLSFLYLYNNOLSG-----SIPEIG 257
QY 594 ---YL-----FDFEHLPN---CASALDFIKLDF---YCGAMASWEKAEPTGIIHNEA 639
Db 258 EICYLRSFLYLDSEALNGSIPASIGNMMNNLSFLYLYNNOLSG---SIPEIG----- 308
QY 640 PEYIYRSRAVSLPFNNKQEFRTLEVTLRDFSKLNKODITYLGKIFFSATSRLR:JIRKAG 699
Db 309 ---YL-----KSLNVLGISENALN----- 324
QY 700 VAGSLSLVSTCKNTKNTYSIMVEASPLTIEDERHITSVNLKLSIHDLONOPL:GLTDSL 759
Db 325 --GSPASIGNKLNNLSRLNLYNNOLSGSI-----ASLGNLNNLSMLYLYNNOLSG:SIIPASL 379
QY 760 GNKKNLTKLIMDNIRKMEEDAIRKAEGLKMKCLFHLTHLSDIOGMDYIV:---SL 815
Db 380 GNLNLSMLYLYNNOLSGSIIPASIG-NNLNSRLNLYNNOLSGSIPEEIGYLS:LYTLDL 438
QY 816 SSEPQD-----LEFIOVSCCLSANAVK:IAQNLHNLVYKLS:LDLSEN 858
Db 439 SNNISNIFIPASFGNMSNLAFLELYENOLASS-----VPEEIGYLSRLNLYNNOLSG 489
QY 859 YLEKDGHEALHELIDRMNVLEQTLALMLPWGDOGSISLKLHEVPOLVKT:ILKNWR 918
Db 490 AL-----NGSIPASFGNMMNNLSRLN-----VNNOLSGSIPEEIGYLSRLNLYNNOLSG 953
QY 919 LTPDTEIR-ILGAFEGKNPLKNFOQLINAGRYSD-----GWL----- 955
Db 534 LSNALNGSIPASFGN-NNLNSRLNLYNNOLSGSIPEEIGYLSRLNLYNNOLSG 955
QY 956 -AFMGVFNKOLVFFDESTKEFELPPALVRKLSQVLSKTEFL 997
Db 592 PASLGNLNNLSMLYLYNNOLSGSIPEEIGYLSRLNLYNNOLSG 628

RESULT 8
US-09-180-439-8
Sequence 8, Application US/09180439
Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
APPLICANT: Hatzixanthis, Kostas
APPLICANT: Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180,439
EARLIER FILING DATE: 1998-12-06
EARLIER APPLICATION NUMBER: PCT/CB97/01249
EARLIER FILING DATE: 1997-05-08
EARLIER APPLICATION NUMBER: GB 9609681.3
EARLIER FILING DATE: 1996-05-09
EARLIER APPLICATION NUMBER: GB 9619924.5
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1016
TYPE: PRT
ORGANISM: Lycopersicon esculentum

Query Match 3.1%; Score 164; DP 4; Length 1016;
Best Local Similarity 22.1%; Pred. No. 7e-06;

Matches 147; Conservative 115; Mismatches 222; Indels 182; Gaps 38;
QY 459 LYSHEPEEVKNGYLGQWVSI-DITSTYSSLRITCGSSVATRAVMKHLAAVYOHG 517
Db 203 LSGFPEI-----GYLRSVTKLSIDIN-----FLSOSIRASGNNLSFLYLYNN 249
QY 518 ILGLSIAR-----PILWR-----QESIOSVKTTRDEI----- 545
Db 250 QLSGSIPEEIGYLSRLKSLIGNIFLSGSIIPASIGNMMNNLSLIDLNNKLSIPEEIGY 309
QY 546 LKAININSFVDCGCIHLYES--TSKSALISOEFAPFOGSKLYINGNIPD---YL----- 595
Db 310 LRSI-----TLDIGENALNGSIPASIGNMMNNLSFLYLYNNOLSG---SIPEIGYLSRLTY 363
QY 596 FDFEHLPN---CASALDFIKLDFYCGAMASWEKAEPTGGIHM-----EAPET 642
Db 364 LDIGENALNGSIPASIGNMMNNLSRLDLYNNKLSG---SIPEIGYLSRLTYLDIGENALNG 421
QY 643 YIPSRVSLPFNNKQEFRTLEVTLRDFSKLNKODITYLGKIFFSATSRLR:QIRKAGVAG 702
Db 422 SIPASIGN-----NNLFMLYLYNNOLSGSIPEEIGYLSRLTYLYGNN---SLNG 469
QY 703 SLVSTCKNTKNTYSIMVEASPLT---IEDERHITSVNLKLSIHDLONOPL:PGTDSL 759
Db 470 SIPASIGNMMNNLSFLYLYNNOLSGSIPEEIGYLSRLTYLYGNN---SLNG 523
QY 760 GNKKNLTKLIMDNIRKMEEDAIRKAEGLKMKCLFHLTHLSD---IGGMDYIVKSL 815
Db 524 GNLNLSMLYLYNNOLSGSIIPASFGNMMNNLSRLN-----FLSDNDLIGEIFVFCNLT 858
QY 816 SSE-----PC---DDEIOLVSCCLSANAVK:IAQNLHNLVYKLS:LDLSEN 907
Db 576 SLEVLYMSRNNLKRPVQOGLNISDLHIS--WSSNFEIGELPSSISNLTSLIKLIDFGRN 907
QY 859 YLEKD-----GNEALHELIDRMNVLEQTLALMLPWGDOGSISLKLHEVPOLVKT:ILKNWR 918
Db 634 NLEBAPFOEGGNISQVDFDMQN--NKLSG-TLPYNSIGSLSLISLHNGELADEIPRS 690
QY 908 ---QIVKLGILKNNRLTDT-----EIRILGAFGR--NPLK-----NFQOL 942
Db 691 LDNCKRLQVLDLGDNDLNDTFFPMWGLTLPBLRVNLTSNKLHGRSSGAIMPPDIRIT 750
QY 943 NLAGNRYSSDGMIAFGVFNKOLVFFDFTSKR-----FLPDPALY-----RKLSQVY 991
Db 751 DLSRNPQD---LPTSLEHLKGMRTVDMKMBEPSESYDDSVYVYKGLLELTVRL 807
QY 992 SKTFL 997
Db 808 SLXTII 813

RESULT 9
US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J
APPLICANT: RATTNER, JEROME B
TITLE OF INVENTION: NUCLEOTIC ACID ENCODING A
TITLE OF INVENTION: TRANSFERRABLE EXPRESSED KINETOCHORE PROTEIN,
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS


```

; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

```

```

Query Match 2.8%; Score 149.5; DB 5; Length 3248;
Best Local Similarity 18.4%; Pred. No. 0.001;
Matches 213; Conservative 195; Mismatches 461; Indel: 287; Gaps 52;

```

```

OY 39 NICEKVDADARGIIMILKSGS-----NLFKSLKEMNPIFODINGQ-- 87
DB 684 NYLDSKSVSEVETOK-LAYMELQOKAFESDQKHQKEIENMCLKT-----SOLTGQVE 733
OY 88 SLFH--QTSEGDLDLAODKLDLHTPSFNFYPLGEDIDIDIFN-----LSTFTEPLV 139
DB 734 DLEHKQLLSTNEIMDKRCYQDLHAEYSLRDLKSKDASLVTNEDHQSL.AFDQOPAM 793
OY 140 WRKDOHHNVEQDITLGLQALQSPCTIGESGKSTLLO-RIMAMWMSGI CKALTKFK 198
DB 794 -----HSPANITIGEGSMSPSESECRLEADQSPKNSAILQNNVDSL----- 835
OY 199 FVPEFLSLSAQGLPFTLDLDIDPGTIRKQTFMAMILKLRQVLFLLDGYKEKPKMC 258
DB 836 -EESLESOKOMNSDLOKCEELVOIKETLEEN-----LKKRQM-----HSPVAETS 882
OY 259 PEPEALIKENHKFNMYITTTTECLRHTRQGCALTAEGDMTEGSAQALIRV-LIKEL 317
DB 883 ORISKOEDPTSAHQ-VAMETLSALENKEKELQDLNDKY-----EEOQAEIQLKKSNNH 936
OY 318 AEGULQIQKSKRLNMLTPFLVYVTCALOMGESEFHSHTQ--TLFHTFYL----- 369
DB 937 LEBSLKELOLSTLSTLEKEMSSITIS---LKKREIEILQENGLKELINSLNOEKNN 992
OY 370 LIOKKN-----HKHGYA-ASDFIIS-----LDHCGDL--ALCEVSFKED TELQDVS 414
DB 993 LIOKSESPANYDERKRSISELSDQYKOEKLLILORCEEGNAVEPL-SQKX-AAQEN 1050
OY 415 SYNEDEVLLTGLCKYTAQRFKPKYKFKFSQOEY-----TAGRRL 455
DB 1051 SKLECLINCTSLCENRKNELDOLKEAFKHEQEFJTLKLAEBERNOMLLELTYOAL 1110
OY 456 SSLTSHREPEEYTKNGYLQKWSISD---ITSTYSSLLRYTCGSSVEATRAV AKHLA 511
DB 1111 RSEMTDNONNSKSEAGLQKOEIMTLKEBONKMKOEYNDILOEN-----DOLMKVIR---- 1161
OY 512 VYOHGCLGLSLAKRPLMRQESIOS--VNTEEOELIKAININSEVEGCIHLYOIS-TSK 568
DB 1162 -RKHEC-----QNISEPIINSVYERESERNOCFKPOMLEVEKEI SLDSY 1206
OY 569 SALSQFEAFQOKSLIYNSGNIPDYLDFEFELPFCAS-ALDFIKLDYFGGMAWMEA 627
DB 1207 NMOVLQELMRKELKLOES-----EREKCLQHELOTITGDLTETSLQ MQ-- 1254
OY 628 AEDTGIHME--APETIYPS-RAVSLFFNMKQEFK-TEVTLRDSKSKNKQOIT LGKI 683
DB 1255 SQEISLQKCEIDAEKRTISGPRELSTSONDNHLOCISQITNNKLNLEKEICEII QAEK 1314
OY 684 FSSATSLRLQIRKAGVAGSLSLVSTCKNIYSLAMEASPL--TIED----- 728
DB 1315 YELVTELENSRSCTITATRKMAEVEKLLNEVYITLMDSGLLGELVEDIPGGEFSQOPN 1374
OY 729 ERHITSVNL-----KTLSHDLQNRPLPGCLTDSLGNLKNLTKLMD----- 771
DB 1375 EQHPVSLAPIDESNSYETJLSDEKVOHMFALQEKFLSLQSEHKILHQQHCOMSSIMSE 1434
OY 772 -----NIKMNEDAIR-----LAEGL-KNLKMKLFLHLSLDTIG EGM 808
DB 1435 LQTYVDSLAKENLVLTSLNRFQGLVKEWQGLGELGLVPSLSSSCVPDSSSLSLGDS 1494

```

```

OY 809 DY-----IVKSLSEPCDLEIQLVSCCLSANAYK-----IIAONLHN-- 846
DB 1495 FYRALLEOTGDMSLLSLNEGAVASNOQSVDEVFSSLDQTYVDSLAKENLVLTSLNRFQ 1534
OY 847 -LVKSLTDLSENVL-----EKDNEALHELIDRNNVLEQTLALMLP 887
DB 1555 DLVKEMQGLGLEGLVPSLSSSCVPDSSSLSLGDSFFYRALLEOTGDMSLLSLNEGAVA 1614
OY 888 WCGDVQSGSLSLKH-----LEEVQVLKGLKNNWRLDTETIRITLGAFFGNLKNFQOLN 943
DB 1615 NOQSVDEVFSSLOEENLTKETTPAPAKGYE-----ELESICEYR-----QSL 1660
OY 944 LAGNVSSDGMWLAFGVFNELKQLVF-----PDFSTKEFLPDPALVRKLSQVLSKLTFLQ 998
DB 1661 KLEEMMESQGIKKNKEI-QELEQLLSERQELDCRKQYLSSEN--EOWOQKITSVLEM 1716
OY 999 EARLVGQCFDDDLV 1014
DB 1717 ESKLAERKQTEQLSU 1732

```

```

RESULT 11
US-08-947-823-3
; Sequence 3, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Jsgouhi
; APPLICANT: Vaghebi, Jafar
; APPLICANT: Bodess, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Confering
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-0702100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-947-823-3

```

Tue Mar 26 11:36:23 2002

us-09-697-089-2.rai

2.7%: score 144; DB 3; length 1255; Gaps 56; Query Match Similarity 18.1%; Pred. No. 0.00067; Indels 452; Mismatches 358; KGSSEC- 66
Matches 216; Conservative 169; Mismatches 358; KGSSEC- 66
28 FVWVNLNEEVNICEVYDQARG---TTHMLK-----KGSSEC- 66
191 FVWVNLNEEVNICEVYDQARG---TTHMLK-----KGSSEC- 66
67 NLFKS-----LKEWNPFLFODL-----NGOSLFHOTSSEGLDLDLADKXLY 109
251 VGLFKOLLETSPDLIREVLPDQEHVTVIPSTGARNHVMWELLLILSDMPKPT 310
110 HPSPLNFY-----PLGSD-----IDIFNLKSTFEPVLM 140
311 HHDKLELDLVGLTREVSTVRLDEEPKKNKGNMOTNCATLIDLENE-LL 363
141 RKDQHNHVEOL-----LNGILOA-10SPCIEGSGKSTL 178
364 KDKLHVYLAALDSSCCFPMSPDGLFMHLHLHNDLSDSNAYSLAKELELYKODL 423
179 LORIAMGSGCKALTKFKEFVFLSLRAGGIFETCDLDT-----223
424 -----KFTRSFV---DAGGLYKDLARKVLDVAYEAKVIDSITVD 463
224 -----PCTIRKOTFAMMLKLRQVYFLDGYNEKPNCEIPEALIKENHFKN 273
464 NGHLHLIFSLPIIK-----IKLKEISALD-----EHPDKGLIYVNSPKP 332
274 MWYVTTTECLRHIFGALTAEGVDMTPSOALIREVLIKEALGIL-1010SKRCL 551
510 VERKSITFDKI-----TVGEFEIN-----ILKRLKISGADLVISITGP 378
333 NLKPTPLFVVI-----TCALMGSESPSHOTLTLEPTFDLIDKRNKH 378
552 GSKTTLAVKYNDKSVSSRDLFRMCTVQDCD-----KRLNLTISQVSDSKLS 605
379 KQVASFIRSLDHCGLALGVSFHKFDELQDVSNEVDLITGLCKYTAORFKR 438
606 ENIDVADKL-----KOLFGRYLVLDVYDWTTPWELT-----640
439 YKFFKSPOETAGRLSLTSHPEEVTNGYLOK-VSISDTTSTYSILRYTGS 497
641 -----RPFESKSGSRI-----LITREKEVALHGLNTDPDLRLRDESELEKRAFGN 693
498 S-----VEATRAVMKH-----LAAYOHCLLGISIAKPLMR--OESLOVAKNTEOE 544
694 ESCPDELIDVGEKEIAENCKGLPLVADLAVIAGRE-KRQVWLEVOSSLSPTLSEVE 752
545 ILKAINISFVCCGHIHVOESTSKALSOFEAFQCKSLYNSGNIPIYLFDFELPN 604
753 VMAVYELS-----SALDFIKLDFYCGANASWEKAEDTGMHEPAPETYPSPASV- 650
605 -----CA-----SALDFIKLDFYCGANASWEKAEDTGMHEPAPETYPSPASV- 650
767 HLPOLLTFASPKDTSITLYELNYFGABGVGKTEMS-----MEEVKVIYMDLIYS 822
651 -----LEFN-----WKEFRTLEVTLRDFSKIKODITYLGIKISANS--LRLOT-----694
823 LVICFNEIGALFOIHDL-VHDCILKARKENLEPDQSSASDLPDQITDDEBEH 881
695 -----KRCGVAGSLSLVSTCKNIYSIMWEASPL--TIDERHITSYNTK 740
882 FGLNFVMDSNKRRHSG-----KHLVSLITIGDQDSDVSAPHLRHLKLV 929
741 LSHDLONORLPGULTSICNLKNTKLIMINIKMNEEDAIKLAGLKNLKMCLF- 917
930 LQD-----TSPTM-----VDSL-----LNEICMLNLR 954
800 HLDIGCMDIYKSLSEPCDLEET-OLVSCCLSAN-AVKILANLHNVKSLIDSE 857
955 YLS-----IDVOKYL-----PLSFSNLMNESLVSSTNRSITVLPRILDLVLRVLSVDA 1006
858 -NLEKDCNEA-----LHLELI-----DRMNVLEOLTRALML-----PWGC 890

DB 1007 CSFPMADSESLIADETKLENTLITELLISYKDFKNIRKPRNOLLSFLKESMDY 1066
891 DVQGSISLILKHEEVPOLYKGLKNMLD-----TEIRIGAFPEKN 934
1067 STQOHWFSLDELETL-ETLSVGFKSNTNDSSVATNRPWDFHPSKTL--WUREF 1123
935 P-----LKNFOOLNAGNRVSDGW-LAFMGVENLKQVFPDFSKER 977
1124 PLTSDSLSTARLPNLELSYHTLHIGEMNGEEDFENIKELFNNOVSIKRW 1178
RESULT 12
US-09-180-439-6 Application US/09180439
Sequence 6, Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: DIXON, Mark S
APPLICANT: Hatzixanthis, Kostas
APPLICANT: Jones, Jonathan DG
APPLICANT: Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
FILE REFERENCE: 620 - 53
CURRENT FILING DATE: 1998-12-06 PCT/GB97/01249
EARLIER FILING DATE: 1997-05-08
EARLIER FILING DATE: 1996-05-09
EARLIER FILING DATE: 1996-09-24
EARLIER FILING DATE: 1996-09-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 799
TYPE: PRT
ORGANISM: Lycopersicon esculentum
US-09-180-439-6
2.7%: score 143.5; DB 4; length 799; Gaps 17;
Query Match Similarity 23.7%; Pred. No. 0.00035; Indels 99; Mismatches 128;
Matches 90; Conservative 62; Mismatches 128; KGSSEC- 66
675 QDITIGKIFSSATSLRDIKRCAGVAGSLVSTCKNIYSIMWEASPL--TIDERHIT 732
209 EELGIL-----RSLTKSLDIN-----FLSGISIPASISGNLNNISFELYKNOLOGSIPEE- 258
733 TSVTNKLTSHDLONORLPGGLTDSIGLKNLNTKLIMINIKMNEEDAIKLAGLKNL 312
259 -IGYLRSLTYLDLKENALNGSIPASISGNLNNISFELYKNOLOGSIPEE- 827
793 MCLFHLHLDIG-----EGMDIYKSLSEPCDLEIOT-----827
313 LTNLHGNNLSLIGLIPASFGNNRNLOALFLNDNMLIOEIPFCVCMILSLELLYPRNNIK 372
828 -VSCCL-----SANVYK--ILANLHNVKSLIDSENYLEK-NEAL 868
373 GKVPQCGNISDLVLSMNSNFGELPSSISNLSLILDFGRNNLEGAIFQCGNIS 917
869 HELIDRANVLEOLTALMLPWGCGDVQGSISILKH-----LEEVQYKGLKRW 917
433 LQVFMON--NKLSG--TLPTNFSIGSLISLHNGLEDELPMSLDCKKLYQDLDGN 489
918 RLTD-----EIRLIGAFCK--NPLK-----NFOOLNAGNRVSDGW 957
490 QLDNTPMNLGTLPELRVLRITSKHLGPIRSSGAETMFPDLRIDLSSNAPSOD--LP 546
958 MGVFENLKQVFPDFSKER 976
547 TSLPEHLKGMKTYDKTME 565

Tue Mar 26 11:36:23 2002

RESULT 13
US-08-514-014-4
Sequence 4, Application US/08514014
Patent No. 5707829
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Carlier, Kerry
TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
NUMBER OF INVENTION: ENCODED THEREBY
CORRESPONDENCE ADDRESS: 12
STREET: 87 Cambridge Institute, Inc.
CITY: Cambridge
STATE: Massachusetts
COUNTRY: Massachusetts
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/514,014
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 498-8224
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 661 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
US-08-514-014-4
Molecule type: Protein

Query Match
Best Local Similarity 2.7%; Score 142; DB 1; Length 661;
Matches 125; Conservative 65; Mismatches 215; Indels 140; Gaps 24;
486 TYSLLRYTCSSVEATRAVYK--HIAAYOHGCLGLSLAKRPL--WROESLOSVMNT 541
DB 72 TTSRLMNT--FDLTRCOIMNHEDTPOSHQSLVLTGNPLIEMAEISLNGPKS-- 126
OY 542 BOEILKAININSVEGCIHLYOESTKSALSOEF--EAFPOGSKLYINGNIDYLPDF 598
DB 127 ---LK-----HFLIOTGISNL--EFLPVNHNLESLSLYSNHI--SIKFE-- 166
OY 599 FEHLPCASALDEFTKIDPYGA-----MASWKA-----AEDGGIHMTEAPET 642
DB 167 ---FDEPPAENLVKDFOMNAIHYSREDRSLDEOMINISLNFNGNNVKIETLAARDST 222
OY 643 YIPS-----RAVSLFPNNKQEFRTLEVLTARDSSKLNKODIT--YLGKLFSSA7SLRLQ 693
DB 223 VFOSLNGGCPNLSVIFNGLONSTOSLWIGTFEDIDEDBSMMLGCLGEMSVETLNLQ 282
OY 694 IKRCAGASL-----SIVSTCKNYSLSIVAS 722
DB 283 EHFSDISTFPCFOTLOELDLATLHLKLPSCMKGLNLTLYVSV--NHFOI-QSIS 340
OY 723 PLTIDENHITSYNNKTLSTIHDIONRLPGCLTSLGMLKMLTGLINDINKNNEEYALK 782
DB 341 AANPSSLTHYINGNKKL-----HIGVGLERLGNLTGLD--LSINDTEASI CCS 389
OY 783 LAGELNKKMKCLPHLTILSDIGEMDYIVKLSSEPC-DLEETOLVSCCLSAANVKTIA 841

us-09-697-089-2.ra1

DB 390 LQ--LKNLSHLOTNLINSHNEPFG-----LQSAFRECQLELDLAFTRHINAFQSPF 441
OY 842 ONLHNYKLSI-----LDLSE-----NYLERKGNALHEILIDPMVNLBOLAL--M 885
DB 442 ONLHNYKLSI-----LDLSE-----NYLERKGNALHEILIDPMVNLBOLAL--M 885
OY 886 LPWGCVOGSSLSLKLHLEVPOLYKLGKMRRLDTEIRITLAFGKNPLKNOOLNLA 945
DB 502 ILSSC---GLSLIDQONFHSLSKMSHVDLSHNSLTDSIDSLSHLKG-----TYLNIA 551
OY 946 GNRVS 950
DB 552 ANSIN 556

RESULT 14
US-08-833-823-4
Sequence 4, Application US/08833823
Patent No. 5969093
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Carlier, Kerry
TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
NUMBER OF INVENTION: ENCODED THEREBY
CORRESPONDENCE ADDRESS: 12
STREET: 87 Cambridge Institute, Inc.
CITY: Cambridge
STATE: Massachusetts
COUNTRY: Massachusetts
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/833,823
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 498-8224
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 661 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
US-08-833-823-4
Molecule type: Protein

Query Match
Best Local Similarity 2.7%; Score 142; DB 2; Length 661;
Matches 125; Conservative 65; Mismatches 215; Indels 140; Gaps 24;
486 TYSLLRYTCSSVEATRAVYK--HIAAYOHGCLGLSLAKRPL--WROESLOSVMNT 541
DB 72 TTSRLMNT--FDLTRCOIMNHEDTPOSHQSLVLTGNPLIEMAEISLNGPKS-- 126
OY 542 BOEILKAININSVEGCIHLYOESTKSALSOEF--EAFPOGSKLYINGNIDYLPDF 598

```

Db 127 -----LK-----HLPLITGICISNL--EFIPVHNLENLESLYLGSNHISSIKF-- 166
QY 599 FEHLPCASALDEIKLDFEYGA-----MASWEKA-----AEDGIGIMEBAPT 642
Db 167 ----PQDPFARNLKYVDFONNMIHYISRDMSLEQAINLSLNFNGNNKGIELGAFDST 222
QY 643 YIPS-----RAVSLEFNNKQEFRTLEVTLRDFSKLNKODIT---YLGKIFSSATSLRQ 693
Db 223 VQOSLNFSGTPIKSVIFNGLQNSTGOSLWLGTFEODIDDDISSAMKLGICEMSVESLNIQ 282
QY 694 IKRCAGVAGSL-----SLVSTCKNITYSLMVEAS 722
Db 283 EHRFSDISSTTFQCFQTOLELDLTATHLKGLPSGMKGLMLKKIYLSV--NHFDQLCQIS 340
QY 723 PLTIEBERITSVNLKTLSTIHDLQNRLPGLUTDSLGNLKLMLTKLIMNTKMNEDAIK 782
Db 341 AANFSLTLYIRGNVKFI-----HLGVGCLLEKLGMLQTLTD--LSHNDIEASDCS 389
QY 783 LAEGLKLNKMKCLFHLTHLSDIGEMDYIVKSLSEPC-DLEIOLVSCSLSANNAVKILA 841
Db 390 LG--LKNLSHLQTLNLSHNEPIG-----LQSOAFKECQQLDLPLRLIIMAPQSPF 441
QY 842 QNLHNLVKIST---LDISE-----NYLEKQNEALHELIDRMNVLEQTLAL--M 885
Db 442 QNLHLEQVNLTYCFDITSNQHLLAGLPVLRHLNLKGNHFDQGTITKTLNLQTVGSLEVL 501
QY 886 LPMGCDVQSGSLSLKHLLEVPQVYKLGKLNKRLDTERIRIGAFGKPKLKNFOQLNIA 945
Db 502 ILSSC---GLSIDQOAFHSLGKMSHVDLSHNSLTGDSIDSLSHLKG-----IYLNIA 551
QY 946 GNRVS 950
Db 552 ANSIN 556

```

```

RESULT 15
US-08-519-547A-6
: Sequence 6, Application US/08519547A
: Patent No. 5994082
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Proteins Essential for the Expression of
: TITLE OF INVENTION: Vertebrate MHC Class II Genes, DNA Sequences Encoding Same
: TITLE OF INVENTION: and Pharmaceutical Compositions
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FISH & NEAVE
: STREET: 1251 AVENUE OF THE AMERICAS
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 10020-1104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WordPerfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/519, 547A
: FILING DATE: 25-AUG-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP94113378.7
: FILING DATE: 26-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: HALEY, JAMES F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: VOS-11
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-596-9000
: TELEFAX: 212-596-9090
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 1130 amino acids
: TYPE: amino acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEITICAL: NO
: ANTI-SENSE: NO
US-08-519-547A-6

Query Match      2.7%; Score 141.5; DB 2; Length 1130;
Best Local Similarity 19.2%; Pred. No. 0.00095;
Matches 168; Conservative 122; Mismatches 335; Indels 251; Gaps 35;

QY 151 QLTGLGLQADSP-----CIIEGESGKSKSLDRIAMWCGSKKALTKRF 199
Db 391 QLAQGLAEVLLAAKHNRRPRETRVIAVLAGAGQCKSWAGAVSRAMACGR--LPQYDF 447
QY 200 VFPL-----RLSRAQGGFLFTDQDLIDPOTIRKOTFMM-----LKLROBVFLL 247
Db 448 VFSVPCHLNRRGDAVG-----LQDLFLSLQPLVAADVFFSHILRRPDRVLIL 498
QY 248 DGVNFKPQNCPEIIBALIKENHREKNNVIYTTTECLRHIFQALTAEGVDMTED--SA 305
Db 499 DAFEELEAOD-----GFLHSTCGPAPAEPCSL 525
QY 306 QALIREVILKEIAEG---LLQIQSKSRLRNIMKT-PLFVYITCAIQNGES----EFHSH 357
Db 536 ROLLAGLFQCKILRCKLTLLTARPRGRVQSLSKADALFELSGFSMEQAQAVMYRKFESS 585
QY 358 TOTLFTHTFYDILLQKNK-----HKHKGVAASDFIRSLDHGQDLALAEVFSHKPFELQD 412
Db 586 GMTF--HODRALTLRKDRPLLSHSH---SPTLCRAVCQSEALLE----- 626
QY 413 VSSVNEVDVLTITGLLCKYT---AQRPKPKYKFFKNSFOETAGRLSSLL---TSH 462
Db 627 ---LGEDAKLPSTLGLVGLGRALDPSPGALAEKLAKAMELGRHRRQSTLQEDQFSA 683
QY 463 EPEEYTKNGVLOKKNVISDITSTYSS--LRLYTCSSVEATRAVVK-----HLA----- 510
Db 684 DVRTWAMAKGLVQHPRAAESELPSPFLQCFGLALMALSLGELKKELPOLYALALPRK 743
QY 511 -----AVYOH-----CCLGLSTAKRPLRROESLOVKKWTTDEITKA 548
Db 744 KRPYNNMLEGVPRLAGLIFQPPARCGLALGPSAASVDRKQVLAARYLKRLOPGLIRA 803
QY 549 ININSFVEC-----GI--HLYOESTSK-----SALSOEFAPFGKSLYI 586
Db 804 ROLLELLCAHBAEBAEAGIWOHVQDELPGRLSFLGTRLPDPDAHVLGKALAEAGQDFSLD 863
QY 587 NSGNIPDYLDFEFELPNCASAL-DFIKLDFYGGAMASWEKAABDTGFI--HMEAPETY 643
Db 864 RSTGI-----CPSGLGSLVGL---SCVTFRRALSDTVALWESLKHQHCFTK 906
QY 644 IPSRAVSLEFNNKQEFRTLEVTLRDFSKLNKODITYIKLITSSATSLSLQIKRCAG--VA 701
Db 907 LLQAAEEKF-----TIEFFRAKSLKDVEDLQKL-----VQTQTRSSSEPTA 948
QY 702 GSLSLVSTCKNITYSLMVEASPLTED--ERHITSVYNLKTLSIHDLQNR-----PGGLT 756
Db 949 GELPAVRDLKLLEFALGPGVSGQAPPKLVRLITLTAASSIQHLDLALSNKIGDEGVQSL 1008
QY 757 DSLGNLKLMLTKLIMNTKMNEDAIKLAEGKLNKMKCLFHLTHLSDIGEMDYIVKSL 816
Db 1009 ATFQPKSLETILNLSQNNITDLAGVAKLAELPSL----- 1042
QY 817 SEPCLEETIOLVSCLSANNAVKIILAQNLHNLVKLSIIDLSENVLEKQNEALHELIDRMN 876
Db 1043 ---AASLLRSLYNNCTICVGAESLARVLPDMVSLRMDVQYNKFTTACAGQDLAASLRCP 1100
QY 877 VLEQTLALMLPMGCDVQSGSLSLKHLLEVPQVYK 912
Db 1101 HVELTAM---WTPITPFSVQ---EHLQOQDSRISL 1129

```

Tue Mar 26 11:36:23 2002

us-09-697-089-2.ral

Search completed: March 25, 2002, 10:55:40
Job time: 40 sec

PI Bertin J, Robison KE;

Neuronal apoptosis
Peptide #4326 enco
Peptide #4339 enco
Peptide #4121 enco
Human CARD-4L prot
Apoptosis related
Human CARD-4L (10m
Murine CARD-4L, pro
Human CARD-4L (10m
Human protein sequ
Human CARD-7 poly
Human NRC beta iso
Human G-protein co
Human NB-ARC and c
Human NRC beta iso
Human NB-ARC and c
Human NRC gamma or
Human NB-ARC and c
Arbidopsis thalia
Arbidopsis thalia
Porphyromonas ging
Porphyromonas ging
Porphyromonas ging
Human CARD-4S part
Human CARD-4S (sho
Human secreted pro
Human AV30₁₋₁₄ pro
Human polypeptide
Human protein sequ
HTM clone 2709055
Human G-protein co
Zea mays Rpg1-2 pr
Apoptosis related

XX MPI: 2001-308628/32.
DR N-PSDB; AAS03945.
XX
XX
XX Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
XX disease -
XX
XX
PS
PS Claim 9; Fig 1; 93pp; English.
XX
XX
XX The sequence represents a human caspase recruitment domain 12 (CARD-12)
CC polypeptide. CARD domains are found in a number of proteins that transmit
CC signals that activate apoptosis and inflammatory pathways in response to
CC stress and other stimuli. Therefore, CARD-12 and its corresponding
CC nucleic acid may be used in treatment and diagnosis of patients suffering
CC from disorders associated with an abnormal level (an increase or a
CC decrease) of apoptotic cell death or abnormal activity of stress-related
CC pathways. The disorders include cancer, viral infections (e.g. caused by
CC poxviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus
CC erythematosus, arthritis), neurological disorders (e.g. Alzheimer's
CC disease, amyotrophic lateral sclerosis), haematologic diseases (e.g.
CC aplastic anaemia, myocardial infarction, stroke), inflammatory and immune
CC system disorders (e.g. Crohn's disease, insulin-dependent diabetes,
CC contact dermatitis, psoriasis, graft rejection), bacterial infections
CC (e.g. tuberculosis, lepromatous leprosy), ischaemic and hypoxic brain
CC injury, kidney ischaemia/reperfusion injury, excitotoxic brain damage,
CC acute bacterial meningitis and liver disease.
XX
XX Sequence 1024 AA;

Query Match	100.0%;	Score 5281;	DB 22;	Length 1024;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1024;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

Db	541	tege k a n n s t v e e g n h y g e s t s a s e f e a f g y s t y n s g n p d y f d f f e	600
Qy	601	h l p n c a s a l d e f k l d e f y g a m a m e k a e d t g i h m e a p e t y f e s r a v s f e f n k o e f r	660
Db	601	h l p n c a s a l d f k l d f y g a m a m e k a e d t g i h m e a p e t y f e s r a v s f e f n k o e f r	660
Qy	661	t l e v t l d f s k n k o d t t y g k f f s a t s r l d o k r c a v a s s t v t c n t y s l a w e	720
Db	661	t l e v t l d f s k n k o d t t y g k f f s a t s r l d o k r c a v a s s t v t c n t y s l a w e	720
Qy	721	a s p t i e d e r i t v t n k t s i h d o n o r p c g l d s g n k n t k i m d i k n e d a	780
Db	721	a s p t i e d e r i t v t n k t s i h d o n o r p g y t c d s g n k n t k i m d i k n e d a	780
Qy	781	t k a e g k n k k w c l f h t l h s i d i g e m d y v k s l s e p c d e e i o l v s c c l a n a v t l	840
Db	781	k l a e g k n k k w c l f h t l h s i d i g e m d y v k s l s e p c d e e i q v s c c l a n a v t l	840
Qy	841	a o n l h n v k s i d l s e n t y e k o g n e a h e i d r m a v e o l t a h a l p w c y o g s t l	900
Db	841	a g n h n v k s i d l s e n t y e k d q n e a h e i d r m v e q t a h a l p w c d y g s t l	900
Qy	901	k h e e v p o l v t g k n n r l d t e r i l g a f f g k n p k n f o o l n g n r r s s p d g n a f w g	960
Db	901	k h e e v p q v k g k n w r l c d t e r i l g a f f g k n p k n f q n n a g r v s s d g w a f w g	960
Qy	961	f e n k o l v f e d f e s t k e f p p a l v r k s o v l s k t f l o a r v l v g o f d d d l s v t g a f	1020
Db	961	f e n k q v f f e f s t k e f p p a l v r k s q s k t f l g a r v y q f d d d l s v t g a f	1020
Qy	1021	l v t a	1024
Db	1021	l v t a	1024

RESULT	2	
AAU02861		
ID	AAU02861	standard; Protein; 1204 AA.
XX		
AC	AAU02861;	
XX		
DT	12-SEP-2001	(first entry)
XX		
DE	Human caspase recruitment domain 12 (CARD-12) polypeptide #2.	
KW	Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;	
KW	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;	
KW	systemic lupus erythematosus; arthritis; neurological disorder; stroke;	
KW	Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;	
KW	aplastic anaemia; myocardial infarction; inflammatory disorder;	
KW	Crohn's disease; insulin-dependent diabetes; contact dermatitis;	
KW	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;	
KW	tuberculosis; ischemic brain injury; hypoxic brain injury;	
KW	kidney ischemia; reperfusion injury; acute bacterial meningitis;	
KW	excitotoxic brain damage; liver disease.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200130971-A2.	
XX		
PD	03-MAY-2001.	
XX		
PPF	26-OCT-2000; 2000WO-US29643.	
XX		
PR	27-OCT-1999; 99US-0161822.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
XX		
PI	Bertin J, Robison KE;	
XX		
WP1	2001-308628/32.	
DR	N-PSDB; AAS03946.	

xx Isolated caspase recruitment domain-12 polypeptide and nucleic acids
 PT encoding them, useful for treating and diagnosing disorders associated
 PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
 PT disease -

Disclosure: Fig 2; 93pp; English.

CC The sequence represents a human caspase recruitment domain 12 (CARD-12)
 CC polypeptide. CARD domains are found in a number of proteins that transmit
 CC signals that activate apoptosis and inflammatory pathways in response to
 CC stress and other stimuli. Therefore, CARD-12 and its corresponding
 CC nucleic acid may be used in treatment and diagnosis of patients suffering
 CC from disorders associated with an abnormal level (an increase or a
 CC decrease) of apoptotic cell death or abnormal activity of stress-related
 CC pathways. The disorders include cancer, viral infections (e.g. caused by
 CC polyomaviruses, adenoviruses), autoimmune disorders (e.g. systemic lupus
 CC erythematosus, arthritis), neurological disorders (e.g. Alzheimer's
 CC disease, amyotrophic lateral sclerosis), haematologic diseases (e.g.
 CC aplastic anaemia, myocardial infarction, stroke), inflammatory and immune
 CC system disorders (e.g. Crohn's disease, insulin-dependent diabetes,
 CC contact dermatitis, psoriasis, graft rejection), bacterial infections
 CC (e.g. tuberculosis, leprosy), ischemic and hypoxic brain
 CC injury, kidney ischaemia/reperfusion injury, excitotoxic brain damage,
 CC acute bacterial meningitis and liver disease.

xx Sequence 1204 AA:

Query Match 98.7%; Score 5214; DB 22; Length 1204;
 Best Local Similarity 98.1%; Pred. No. 0;
 Matches 1014; Conservative 1; Mismatches 1; Indels 18; Gaps 1;

QY 1 MNFTKDNRAIIQRMGMVIVQITDDLFVWVNLNREEVNIICCEVEDDARGITIMILK 60
 Db :|||||
 Db 140 vnfikdnstraiqrmgmvtvkiqitddlfvwnvlnreevniicceveddaargitihmlik 199
 QY 61 KGSSECNFLKSLKEMNYPFLQDLNGOS-----LFHQTSEGLDDLA 102
 Db kgsescnflkslkemnyplfqdlngosfeetqmwfnfntssliglthqsegdllda 259
 QY 103 QDLKDLVHTPSFLNFPYIGEDIDIIFNLSKSTFTEPVLMRKDOHHHRYBQLTLNGLLAQ 162
 Db :|||||
 Db 260 qdlkdlvhtpsflnfpdyigedidifnlkstftepvlwrkqhhhrveqlclnglqldq 319
 QY 163 SPCTIEGSSGKSTLQRIAMLMGSKCKALTKRFKFFRFLRLSAOGLEFETLCDDLLD 222
 Db :|||||
 Db 320 spctiegsygkstllqrlamlwsgkckalktkfvflrlsraqgylfctldcqlld 379
 QY 223 IPGTIRKOTFAMLLKLRQRLFLLDGYNFEPKPNCPFEIALIKENHFKNMVITTTTE 282
 Db :|||||
 Db 380 ipgtirktfamlkirkgrvllldgynefkpnpcpeialikenhfkmvlttcte 439
 QY 283 CLRHIFQFALTAEVGMTEDSAQALIREVLKELAEGLLQIQSRCLNLMKTPLEFV 342
 Db :|||||
 Db 440 clrhifqfaltaevgmtdesagallirevllikeaeglllqiksrclnmktplfv 499
 QY 343 ITCAIOMGESEFHSHTQTLTFHFYDILLQKNKHKKGVASDFRSLRDHGDALAEVF 402
 Db :|||||
 Db 500 itcaiomgesefhshtqtlthrfydlilqknkhkvaaadfrslrdhgdalaevgf 559
 QY 403 SHKDFELQDVSVNEDVLTGGLCKTAAORFKPKYFFKFSFOEYTAGRRLSSLLTSH 462
 Db :|||||
 Db 560 shkdfelqdvsvnedvlttgllckytaqrfrpkkyffkfsfgeytaqrtrisslltsh 619
 QY 463 EPEEYTKNGYLGQKMSISDITSTYSSSLRYTCSSVEATRAVMKHLAAVYQHCGLLGLS 522
 Db :|||||
 Db 620 epeeetkngyylgkmvsvsditstysllrytcgssveatravmkhlaavyqhgcllgls 679
 QY 523 IAKRPIMROESLOVKNTEDEILKATININSFVCGTHLYQESTSKSLSOEFEPFGK 582
 Db :|||||
 Db 680 iakrplwrgeslgvknctedeilkatinsfvcgthlyqestskslsoefefgk 739

QY 583 SLVINSGNIPDYLFDFEFELPNCASALDFIKLDFYGGAMASWEKAEDGCIHMEAPET 642
 Db :|||||
 Db 740 slvinsgnipdyldfdehlpncasaldfikldfyggamaswekaedcgdlhmeapet 799
 QY 643 YIPSRVSLFFEMKQDFRTLLEVTLRDFSKLNKQDITLYGKIFSSATSLRLQIKRCAGVAG 702
 Db :|||||
 Db 800 yipsrvslffemkqdfrtllevtlrdfskslnkqditylgkifssatsrlqikrcagvag 859
 QY 703 SISLVLSCKNTISLMVEASPLTIEDERHITSVNLKTLSHDLQNLPGGLDLSGLNL 762
 Db :|||||
 Db 860 sislvlsckntislmveaspltietheritsvnlktslshdlqnglpggltdslgnl 919
 QY 763 KNLTLIMDNIRKMNEDEIKLAEGLNKKKCLPHLTLSIDGEGMDYIVKSLSEPCDL 822
 Db :|||||
 Db 920 knltlimdnirkmedeiklaeglnkkcclphltlsidgegmdyivkslsepcll 979
 QY 823 BEIQIOLVSCLSANAVKILIAONLHNLVLSIIDLSENYLEKDGNEALHILIRMNVEQLT 882
 Db :|||||
 Db 980 eeiqlvscclsanavkilaqnlhnlvlsiidlsenylekdgnealhelidrmvleqlt 1039
 QY 883 ALMLPMGCDVCGSSLSLKLHEVQVLVGLKMKRLTDEIRILGAFFGKNPLKNFOOL 942
 Db :|||||
 Db 1040 almlpmgcdvgsslsllklhevpqlvlglnkwrldtelirrigatfgknplknfql 1099
 QY 943 NLAGNRVSSDGLAFMGVEENLKOLVFPDFSTKEFLPPALVVRKLSOVLSTFLQEARL 1002
 Db :|||||
 Db 1100 nlagnrsvsdglafmgveenlkolvfpdfstkeflppalvvrklsovlstflqearl 1159
 QY 1003 VGMQFDDDDLSVIT 1016
 Db :|||||
 Db 1160 vgmqfdddddsvit 1173

RESULT 3

AAM23595 AAM23595 standard; Protein; 841 AA.

AC AAM23595;

DT 12-OCT-2001 (first entry)

DE Murine EST encoded protein seq ID NO: 1120.

KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KW diagnostics; forensic test; gene mapping; genetic disorder;

KW biodiversity; gene therapy; nutrition.

OS Mus musculus.

PN WO200154477-A2.

PD 02-AUG-2001.

PE 25-JAN-2001; 2001WO-US02687.

PR 25-JAN-2000; 2000US-0491404.

PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

PI WPI: 2001-476164/51.

DR N-PSDB; AAB98254.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use -

XX Claim 20; Page 832-833; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomlato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC biologically active sequence.
XX

50 Sequence 841 AA;

Query Match	81.4%	Score 4297	DB 22	Length 841
Best Local Similarity	99.6%	Pred. No. 0		
Matches 838	Conservative 0	Mismatches 3	Indels 0	Gaps 0

QY	184	MLMSSGCKALTFRKFEFFPLFSLSAOGLEFETLCDDLDLPGTRKOTPFAMMLIKLRORV	243
Db	1	mlwsgckaltkfkfyfifrlstraggfletlcoqldlprglctkqclmamlkrltrqv	60
QY	244	LFLDDYNEFKPQNCPELTALIKENHBFKMNVIVTTTTTECLRHRIOFGALTAEGWMTED	3033
Db	61	lflldgynefkrpncpeleaalkenhbfkmnvivtttteeclrhrgfaltaeygwmted	120
QY	304	SAOALIREVLKLELAEGLLIOIOKSCRLRMIMKTRPLRVTVTCALOMSESFHSHTOOTLE	3633
Db	121	sagelirevlkelaeglllqdgstcrlrmkprlvttvltcaqlmgesehshctcltf	180
QY	364	HTFVYDLLIOKNKHNKGVAAADFRSLRCDHOGDIALCEGVFSHKRDEFLODVSVNEDVLLT	4233
Db	181	htfvdlllqknknhkbgvaadfrslrldhcgylalegvfshktdfelqdvsvnedvllt	240
QY	424	TGLLCXYTAQOFRRKRYFHFHKSFPDEYTAGRRSLSLTSHPEYVTKNGTLOKMNVSIDI	4833
Db	241	tglclxytaqofrkryfhfksfdeytaagrrslsltshpeevetkngylqkmnvsidi	300
QY	484	TSTYSSLIRYCGSSSVATRAMKHLAAVYOHGCLLELSTAKRPLMROESLOSVKNTTEQ	5433
Db	301	tslyssllrycgssvvaetramkhlaavyohgcllglstakrplwrgeslgsvkntteq	360
QY	544	EILKAININSFVECGIHLYOESTSKSALSOFEAFFOQKSLYINSGNIIPDYLEDFEENLP	6033
Db	361	eilkaininsfveegihlygestksalsqfeafifgkslyinsgnipyldlfehenp	420
QY	604	NCASALDPRITKIDFNGMAMSEKAEADONGIHMEAEETVYPSRNVSLPFNMKOEPRFTE	6633
Db	421	ncasaldpritkldfngmamaswekaeetdvgihmeaeetyprsnvslpfnmkoeprfte	480
QY	664	VTLRDPSKLNKODTIVYLGKIFSSATSRLQIKRCAGVAGSLVLSYCKNIYISLMEVAP	7233
Db	481	vtrldpsklnkodytvgkifssatsrllqikrcagvagslsvlscnkniylslmevasp	540
QY	724	LTFIEDERITSVTMLKTLSTHDLONORLPGLDLSCLNKLNLTKLINDNIKMNEDATKL	7833
Db	541	ltfederitstvtmlkltsthdlonorlpglldslcnlknlnltklnindnikmnedaikl	600
QY	784	AEGKLNKLKMKLPHLTHLSDIGEMDITVSLSEPPDLEIOIVSCCLSANVKTIAON	8433
Db	601	aegklnklkmcplhlthlstdigemdytvsllseppdleaivscclsanaavkllaqn	660
QY	844	LHNLVKTSLDLSBNYTEKDGENALHBLIDRMNVLEOLUTALMLPWGDVCGSLSLTKHL	9033
Db	661	lhnlvktslidlslsenytekdgnealhelidrmnvleoltalmlpwgdvgsslsllkhl	720
QY	904	EEVYQVLYKLGLKNNRLDTERIRLIGAFPGKNPLKKNFOQLMAGNRVSSDCMLAPKGVFEN	9633
Db	721	eevpqlyvklglknnrltdterirlligafifgknplknfqqlnlaqrvssdgwlaftmgyfen	780
QY	964	LKOLVFPFSEFKELPPPALVRKLSOVLSTKTFLOEARLWGWOFDDDDLSVITGAFPLVY	1022
Db	781	lkqlvfpfsefkelpppalvyrklsqvlstktfifgearlvwgwfddddsvitgafivc	840

QY	1024	A	1024
		-	
Db	841	a	841

RESULT 4

ID	AA	256	40	standard;	Protein;	642	AA.
AAM25640							

AC AAM25640;

DT 16-OCT-2001 (first entry)
 YV

DE	Human protein sequence	SEQ ID NO:1155
XY		

KM Human; ulcer; HIV infection; human immunodeficiency virus;
KM antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KM antibacterial; endocrine; cardiac; central nervous system; vitruide;
KM anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KM antiagregant; haemostatic; vulnery; antitumor; osteopathic; eczema;
KM dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KM neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KM immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KM antiaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KM cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KM genetic disease; haematopoietic disorder; platelet disorder; asthma;
KM thrombocytopenia; osteoporosis; severe combined immunodeficiency;
KM allergic rhinitis; diabetes; multiple sclerosis; depression;
KM Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KM neurological disorder.

05 Homo sapiens.

PN W0200153455-A2.
XX

PD 26-JUL-2001
xy

22-DEC-2000; 2000WO-US35017.

PR 21-JAN-2000: 2000US-0488725

XX 43-APR-2000; 200005-0522317
XX

XX
XX
XX

(" " " " " ")
" " " " " "

XX
XX
XX

XX	1	4	3	2	1
XX	1	4	3	2	1

DR N-PSDB; AAH99581.

PT Isolated human polyr

XX

XX
XX

CC AAM25963. The proteins can have act

CC antiarthritic; immunosuppressive; antibacterial endocrine; cardiac;
CC central nervous system; virocid; anti-HIV; fungicide; antimetagen;
CC cardiovascular; antihaemic; antiaggregant; haemostatic; vulnerary;
CC anticure; osteopathic; dermatological; antiallergic; antasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic


```

Db 799 pstkagpkivshllhvdn-----keslenisenddylkhpelslqmqj 843
QY 546 LKAI-----NINSFVEGCIHL-----YOESTSKALSQEFFAFGKSLYINSGNIPD 593
Db 844 lrglwqicpqayfemvsehllvialktaqgnt-vaacspfvigfqlgrtlitlganl-q 901
QY 594 YLPD-----FFEHLPNC--ASALDFIKIDFYGC--AMASWE 625
Db 902 yfifhpeslslrsihfrlgnktsprahsvletcdksqvpldqdyasafemewe 961
QY 626 -----KAEDTGTGIMEAPETY-IPSRASVLFPMKQEFRTLEVTL 666
Db 962 rnlakeednvksymdqmrgrspdlstgywklspkqykip-----cleadv 1006
QY 667 RDFSCLKNKODITYLGKIFSSATSRLQIKRCAGVAGSLSVLSTCK-NIYSLWEASPLT 725
Db 1007 ndldvvgqdmlelmtvfassqrlhlnhsrgfiesirpalelskavckcsklskls 1066
QY 726 IEDERHITSVTNLKTLSIH--DLQNRPLPGGLTDSLGKLNKLTLMNDIKM----- 775
Db 1067 aaeqellltplsleslsvsgtlsgdqdlfpn-ldkftckelsvdlagnlnvfvspiee 1124
QY 776 -----NEEDATKLAELGKLNKMKLPFLH--THLSDIGEMDIYKSLSE 818
Db 1125 ffnfhmeklllqisaeaydpskivklignspnlvfhkcnfisdqslmtulvs----- 1179
QY 819 PC-DLEBIQVSCCLSANAVKILQNLNLVLTSLDL-SENYLEKQNEALHELDIDRMN 876
Db 1180 -cklteiktsdftf-qavpfa-slpnfisklhlnlegqfpedetsekfaylsgls 1235
QY 877 VLEDTLMLPMWCGDVGSLSSLLKLEBPQVLKLGKMKRLTDT-----EIRILCAFF 931
Db 1236 nlee--lllptgdqlyrvaklllqgcqqlhcltrvlstfklindsvvelakvalsqgf- 1291
QY 932 GRNKLKNOOILNAGN-RVSSDGLAFMGVGENLKOLVFPDFS--TKRFLPDALVRKL 987
Db 1292 -----qlenlksinhkileegrynfqaldmnpnlqeldisrhfecikagatlvksl 1346
QY 988 SQVLSKLTFLOEARLVGMQFDDDLVSVI 1015
Db 1347 sqcvlrlprlrlrmlslwldaddtall 1374

```

RESULT 6
ID AAY14079 standard; Protein: 1403 AA.
XX AAY14079;
AC
DT 20-JUL-1999 (first entry)
XX
DE Gonadotropic hormone protein sequence.
XX Gonadotropic hormone; excessive ovulation animal; transgenic animal;
XX totipotent cell; somatic cell chromosome.
OS Homo sapiens.
XX JP1113444-A.
PN
XX
PD 27-APR-1999.
XX
PF 14-OCT-1997; 97JP-0280830.
XX
PR 14-OCT-1997; 97JP-0280830.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (SAKA-) SAKAI H.
XX MPI; 1999-320709/27.
DR N-PSDB; AAX58000.
XX
PT An excessive ovulation animal - useful for improving the

```

PT productivity of animals
XX
PS Claim 3: Page 11-14; 18pp; Japanese.
XX
CC This sequence represents a gonadotropic hormone.
CC The invention relates to an excessive ovulation animal, which is a
CC transgenic animal with a totipotent cell containing a DNA fragment
CC containing a promoter sequence and a gonadotropic hormone coding
CC sequence. The DNA fragment is in the somatic cell chromosome. The
CC excessive ovulation animal is useful for improving the productivity of
CC animals. The method can improve the productivity of a useful animal.
SQ Sequence 1403 AA:

```

Query Match 12.4%; Score 653; DB 20; Length 1403;
Best Local Similarity 23.7%; Pred. No. 9, 1e-46;
Matches 263; Conservative 197; Mismatches 410; Indels 238; Gaps 42;

```

QY 69 FLKSLKEMNPLRFODLNGQ-----SLFHQTSQGLDD----- 100
Db 344 flgmks-saeavrpdlqsrgecellletsenledslavgrplvpmagaeqwfgeakn 402
QY 101 LAODKLDLYTPSFLN-----YPLGEDIDITNLSKSTFEPLYMKKDDHNR 148
Db 403 lneqlraaytsasfrlmslldssdlatdhlldgdisl---aaskhskpy----- 449
QY 149 VEOLTRGLQALQSPICIEGESGKSTLQRLAMLMGSKCKALTKFVFELRLS-- 206
Db 450 qeplvprevglnhsvmcvgeagsgkvtllkklafiwascoplntfqlvlylsist 509
QY 207 RAOGLEFETLDDLLDIPGTIRKOTFMAMLLKLQRYVFLLDGNEF--KPQNCPEIAL 264
Db 510 rpedglaslcdqllkeqsvtemcmrnllyqkngylfllddykelcsipq--vlgkl 566
QY 265 IKENHRKKNWVITYTTTECLRHTRQFCALTAEVGDMEDBSAQLIRVILKELA--EGLL 322
Db 567 lqknhstcllavlfnrardirylellelkafrlytvclrlfshmtlrjtkim 626
QY 323 LQIQSRCLNLMKTPLEVVITCAIQGESEFHSHTQTFEHTFYDLLIQNKKHKNHGA 382
Db 627 vyfgkngslqkqktrflvaicahwfygfpdsfdvavkysymelstlrnk-----a 680
QY 383 ASDFIR-SLDHCGDLEAGVFSHKFDELDV--SSVNEVDVLTGLCKVYTAQREPKY 439
Db 681 taellkatvsscgealqkfjccofendddlaeaydededltmcmsftagrlrfy 740
QY 440 KFFHKSFOETTAGRRLSSLTSHPEEVTGNGYLOKMSISDITSYSSLLRYTCSSV 459
Db 741 flspalqetlagnrllellsdrgqbdlqlynlkqinspmmtvsaynflnyv--ssl 798
QY 500 EATRA--VMKHLAAYVQHGCCLGLSLAKRPLMWQESLQSVKN-----TTEQEI 545
Db 799 pstkagpkivshllhvdn-----keslenisenddylkhpelslqmqj 843
QY 546 LKAI-----NINSFVEGCIHL-----YOESTSKALSQEFFAFGKSLYINSGNIPD 593
Db 844 lrglwqicpqayfemvsehllvialktaqgnt-vaacspfvigfqlgrtlitlganl-q 901
QY 594 YLPD-----FFEHLPNC--ASALDFIKIDFYGC--AMASWE 625
Db 902 yfifhpeslslrsihfrlgnktsprahsvletcdksqvpldqdyasafemewe 961
QY 626 -----KAEDTGTGIMEAPETY-IPSRASVLFPMKQEFRTLEVTL 666
Db 962 rnlakeednvksymdqmrgrspdlstgywklspkqykip-----cleadv 1006
QY 667 RDFSCLKNKODITYLGKIFSSATSRLQIKRCAGVAGSLSVLSTCK-NIYSLWEASPLT 725
Db 1007 ndldvvgqdmlelmtvfassqrlhlnhsrgfiesirpalelskavckcsklskls 1066
QY 726 IEDERHITSVTNLKTLSIH--DLQNRPLPGGLTDSLGKLNKLTLMNDIKM----- 775

```

Db 1067 aaegeillitlpleslevsgtqsgdqlfpr--ldkflclkelvsydegnlnvsvipdee 1124
QY 776 -----NEEDAIKLAEGLNKKMCLEFHL--THLSDIGCMYIVKSLSE 818
Db 1125 fpuflhmeklllqsaeydpsklvklqnsplnhvfhkcnffsfqslmtmlvs----- 1179
QY 819 PC-DLEEIOLVSCCSAANAUKILAOINLHVKSLTLDL-SENYLEKDCNEALHELIDRMN 876
Db 1180 -cklteikfsdfff-qavpfva-slpnfisikllnlegqgfpoeeekfayllgsls 1235
QY 877 VLEOLTALMPWCGDVQGSLSLKHLEEVPOLYVKGKMMRLDPT---EIRILGAF 931
Db 1236 nlee---lllptgdgiyrvaklllqgcqqlnclyvlstffkclndssvveakvaissgff- 1291
QY 932 GKNPLKKNQOQLNAGN-RVSSDGMALAFMGVFNENLKQLYFPDES---TKEFLPDPALVKKL 987
Db 1292 ----qklenklklsinhkltteegynrfqaldmnpnlqelldisrftccikagatvksl 1346
QY 988 SQVLSKLTFLQEARLVGQMFDDDLSTVI 1015
Db 1347 sqcvlrlprllrlmnlswladdaiall 1374
RESULT 7
AA09539
ID AA09539 standard; Protein; 1403 AA.
XX
AC AA09539;
XX
DT 20-JUL-1999 (first entry)
XX
DE Human apoptosis inhibiting protein #1.
XX
KW Human; apoptosis inhibitory protein; apoptotic disease; diagnosis;
KM spinal muscular atrophy.
XX
OS Homo sapiens.
XX
PN JP11116599-A.
XX
PD 27-APR-1999.
XX
PF 14-OCT-1997; 97JP-0280831.
XX
PE 14-OCT-1997; 97JP-0280831.
XX
PS 14-OCT-1997; 97JP-0280831.
XX
PT (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI: 1999-323531/27.
XX
N-PSDB; AAX56272.
XX
PT New apoptosis inhibitory protein - useful for determining mechanism
PT of various apoptotic diseases e.g. human spinal muscular atrophy
XX
PS Claim 1; Page 4-8; 16pp; Japanese.
XX
CC The present sequence represents a human apoptosis inhibitory protein.
CC The apoptosis inhibitory protein is useful for the elucidation of
CC the mechanism of various apoptosis diseases such as human spinal
CC muscular atrophy and the diagnosis, the prevention and the treatment
CC of such diseases.
XX
SQ Sequence 1403 AA:
Query Match 12.4%; Score 653; DB 20; Length 1403;
Best Local Similarity 23.7%; Pred. No. 9,1e-46;
Matches 263; Conservative 197; Mismatches 410; Indels 238; Gaps 42;
QY 69 FLKSLKENNYPLFPDDLNGQ---SLFHOTSGDLDL----- 100
Db 344 flgmks-saavtpdlgrgcllettsesnldestiavgiyvpemagagwfgaekn 402

QY 101 LAODCKDLYHPPSFLNF-----YPLGEDIDILFNLKSTFTEPLVMRKQDHNHR 148
Db 403 lneqrlraaytsesfrhmslldisclatdhllgclidsl-----askhskpv----- 449
QY 149 VEOLTLNGLLOALOSPCEIEGSGKSTLQRIAMLMGSGKALTFKFEYFRLS-- 206
Db 450 qeplvlpevfnglnsvmevegeagsgkvlkklafvwasgcpblmflqvlvylsist 509
QY 207 RAQGLFETLCDQLDITGTIRKQTFMAMLLKROVYFLDUGYEF--KRONCPETAL 264
Db 510 rpedlaelidcdqllekegsyvelemcmnlilqklngvifliddykelsipq---vialk 566
QY 265 IKENRFRKMNVTVTTCCLRHIIROFALTLAEVGOMTEDSAQALREVLIKELA--EGLL 322
Db 567 lqkmlstclliavtrnarriditryletlieikafpynvcillrkfshmtlrlkf 626
QY 323 LQIQSRCLRLNLMKTPLEFVITCAIQMGSEFHSHTQTLTFHTFYDLLQKNKHKQVA 382
Db 627 vyfgkngslqkikqkpllvaaicahwfgypfdpsfdavafksymernltnk-----a 680
QY 383 ASDPFR-SLDHCGDLALGCVFSHKFDFELQDY--SSVNDVLLTTGLCKYTAQRFKRY 439
Db 681 laelkavsvscgelalqkfiscfcfeinddlaeagvdedlclmclnskftaqrflrpfy 740
QY 440 KEFHKSFOEYTAGRRLSLSLTSHEPEVTKGNGYLOKKWVSTDIRSTYSSLLRYTCSSV 499
Db 741 flspatqelflagmrllellsdqrqehgdqlylqkqlnspmtvsaaynflnyv--ssl 798
QY 500 EATRA---VMKHLAAVYOHGCLLGSLIAKRLPLROESLOSRYK-----TTEOEI 545
Db 799 pstkegpkvlvehllhlvdn-----keslenisenddyllkhpelslqnl 843
QY 546 LKAT---NINSFVDCGTHL-----YQESTKSALSQEFKFAFGCKSLYINSGNIPD 593
Db 844 lrglwgqicpqayfsmvsehllvlalktaygsnt-vaacspflqflgrrtlilgalnl-q 901
QY 594 YLPD-----FPEHLPMC--ASALDFIKLDFYGS--AMASWE 625
Db 902 yffdhpeslslrlsrhfrirgnktsprahfsvletcfdksqyrltdqdyasafermewe 961
QY 626 -----KAEDTGGIMEEAPERY-IPSRVSLFPMWKOEFRTLEYTL 666
Db 962 nlaekednvksymdmqrraspsldstygwlspkqykfp-----cleadv 1006
QY 667 RDFSRLNKQDITTYLCKIRSSATSLRLQIKRCAGVAGSLVLSTCK-NITYSLMVEASPLT 725
Db 1007 ndldvvgqdmlelmtvfasqrlelhnhsrgfieslrpalelskasvctkosisklels 1066
QY 726 IEDERHITSVNLKTLSTH---DLQNRLLPGGLPDSLGNLKNLTKLINDNFKM----- 775
Db 1067 aaegeillitlpleslevsgtqsgdqlfpr--ldkflclkelvsydegnlnvsvipdee 1124
QY 776 -----NEEDAIKLAEGLNKKMCLEFHL--THLSDIGCMYIVKSLSE 818
Db 1125 fpuflhmeklllqsaeydpsklvklqnsplnhvfhkcnffsfqslmtmlvs----- 1179
QY 819 PC-DLEEIOLVSCCSAANAUKILAOINLHVKSLTLDL-SENYLEKDCNEALHELIDRMN 876
Db 1180 -cklteikfsdfff-qavpfva-slpnfisikllnlegqgfpoeeekfayllgsls 1235
QY 877 VLEOLTALMPWCGDVQGSLSLKHLEEVPOLYVKGKMMRLDPT---EIRILGAF 931
Db 1236 nlee---lllptgdgiyrvaklllqgcqqlnclyvlstffkclndssvveakvaissgff- 1291
QY 932 GKNPLKKNQOQLNAGN-RVSSDGMALAFMGVFNENLKQLYFPDES---TKEFLPDPALVKKL 987
Db 1292 ----qklenklklsinhkltteegynrfqaldmnpnlqelldisrftccikagatvksl 1346
QY 988 SQVLSKLTFLQEARLVGQMFDDDLSTVI 1015
Db 1347 sqcvlrlprllrlmnlswladdaiall 1374

AC AAW20033;
XX
XX
DT 06-OCT-1997 (first entry)
XX
DE Neuronal apoptosis inhibitor protein (NAIP).
XX
XX Neuronal apoptosis inhibitor protein; NAIP; diagnosis;
KM therapy; cancer; AIDS; amyotrophic lateral sclerosis;
XX spinal muscular atrophy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 600 /note= "encoded by AAA"
TT Misc-difference 919 /note= "encoded by CCA"
XX
XX
XX MO9726331-A2.
XX
XX 24-JUL-1997.
XX
XX 17-JAN-1997; 97WO-IB00142.
XX
XX 19-JAN-1996; 96GB-0001108.
XX
XX (UYOT-) UNIV OTTAWA.
XX
XX korneluk RG, Mackenzie AE, Robertson G, Roy N, Tamai K;
XX WPI: 1997-385335/35.
XX DR N-PSDB; AAT71266.
XX
XX
XX New neuronal inhibitor of apoptosis - useful for diagnosing and
PT treating, e.g. cancer, AIDS or amyotrophic lateral sclerosis
PS
PS Claim 41; Fig 7A-L; 102pp; English.
XX
XX Novel human neuronal apoptosis inhibitor protein (AAW20033), or NAIP,
CC is a negative regulator of apoptosis, partic. neuronal apoptosis
CC and, when deficient or absent, contributes to neurodegenerative
CC phenotypes such as spinal muscular atrophy (SMA) and amyotrophic
CC lateral sclerosis. Its amino acid sequence was deduced from a
CC cDNA clone (AAT71266) obtd. from a human foetal spinal cord cDNA
CC library. NAIP polypeptides, esp. those containing at least two
CC BIR (baculovirus IAP repeat) domains, can be expressed in host-
CC vector systems and used to increase or induce apoptosis for the
CC treatment of AIDS, neurodegenerative disease, myelodysplastic
CC syndromes or ischaemic injury, to screen for (ant)agonists, or to
CC produce antibodies useful for inhibiting apoptosis.
XX
XX Sequence 1403 AA;
SQ
Query Match 12.3%; Score 652; DB 18; Length 1403;
Best Local Similarity 23.7%; Pred. No. 1.le-45;
Matches 261; Conservative 196; Mismatches 418; Indels 228; Gaps 41;
OY 69 FLKSLKENVNPLFQDLNGO-----SLFPHOTSGDLDL----- 100
DB 344 flgmks-saevtpdlrgelcellettsenledslavpilyemagseqwfgaekn 402
OY 101 LAODLKDLYHPPSFLNF-----YPLGEDIDIFNLKSFTEPVLMRKQDHHNR 148
DB 403 lneqlraaytsstfrimaldsldatclhllgdisl---askhskrv----- 449
OY 149 VEQLTNGLALQSPCTIEGSGKSTLQRIAMLMWSGCKALTFKFEVFLRLS-- 206
DB 450 qeplvlpvfgnlnsvmevegaqsgkvllkklafwasgccpllnfqlvfyfslst 509
OY 207 RAQGLFLELDCQLDIDIGTIRKQTFMAMLLKLRQVFLFDGYNF--KQNCPEIAL 264
DB 510 rpdeglaasfcdqllkeqsvtemcmrnlqqlknqvlllddykelslpg---vlgkl 566

OY 265 IKENHREKMNVTVTTECLRIHQFGALTAEVGMTEDSAQALIREVLKELA--EGL 322
DB 567 lqknhstclliavtrncraridrlrylletllgaqfpylnvclrlkfshmtlrfkm 626
OY 323 LOIQKSRCLRLNLMKTPLEPVITCAIQMGSEFHSHTQTLTFEFDLLQKNKHKCVA 382
DB 627 vyfngnglqkqktrltvaalcawfygpfpsiddvaavfsymertslrnk-----a 680
OY 383 ASDFTIR-SLDHCGDLALGCVFSHKFDFELQDV--SSVNEVDLLTTGLCKYTAQRFKRY 439
DB 681 laelkatvsscgejalqkfiscctefndddlaeagvdeddltnclmskflaqrllrfy 740
OY 440 KFEHKSQFEXTAGRLSSLLTSHPEEYTKNGYIQKNVTSIDITSSTSSLLRTGSSV 499
DB 741 rllspaigeFLAGMLLelldsdqehdqlghlyhklqinspmmtvsaynnflnyv--ssl 798
OY 500 EATRA---VMKHAAYVOHGCILGLSIAKRPLMRQESQSVKN-----TTEOEI 545
DB 799 pstkagpkivshllhvdn-----keslenlenddyllkhpelslmgql 843
OY 546 LKAI---NINSEVECGIHL-----YQSTSKSALQFEAFQOKSLYINSGNIPD 593
DB 844 lrglwqicpqayfsmvsehllvllaklayqsnr-vaacsprvqlgqrltlgalnl-q 901
OY 594 YLFD-----FFEHLPRC--ASALDFIKLDFYGG--AMASWE 625
DB 902 yfddhpesllsrhlsfrgnktsprahfsvletctfcksgvptldqdyasaafepmwe 961
OY 626 -----KAAEDTGIHMEBAPERY-IPSRAVSLFFMKQEFRLLEVTL 666
DB 962 rnlakednvksyndmgrraapdlstgywklsbkypk-----cleadv 1006
OY 667 RDESKLNKODITTYLGGITRSATSLRLQIKRCAGVAGSLVLSTCK-NIYSLMWEASPLT 725
DB 1007 ndidvgqdmlellmtvfaeqrlhlnhsrgflsrlpalelskaavtcsislkels 1066
OY 726 IEDERHITSVNLKTLSTH--DLQNGRLPGGLDLSGLNKLNLKIMDNMKM----- 775
DB 1067 aaeqelllltplsleslevsgtlsgdqqlfpr--ldkfcllelsavdlegnlnvsvipee 1124
OY 776 -----NEEDAKILAEGLNKLKMKCLEFHL--THLSDIGEGMDYIVKSLSE 818
DB 1125 fprfhmeklllqdsaeqdpkylvklinspnlnvfhkcnffdsfgslmtlmsv----- 1179
OY 819 PC-DLEETQLVSCCISANAVKILQNLHNLVKLSTLD--SENYLEKQDNEALHELIDRMN 876
DB 1180 -ckriteikfsdsff--qavpfva-slpnfiskllnlegqqfpedetsekfayllgsls 1235
OY 877 VLEQETLALMPCGDVQGSLSLKLHLEEVQVLKLGAKNRLDTERILIGAFGKNPL 936
DB 1236 nlee---lllptgqlyrvaklllqqcqlnclrlvlsffk-llndsvvelakvaahgff 1291
OY 937 KNFQDLNLAGN-RVSSDGLWLFMGVEFNLLKQDVFFEDS---TKFELPPALVRKLSQVLS 992
DB 1292 qklenklsinhkltteegynrfqaldmmpnlqldslsrhttecklkagatvkslsqcvl 1351
OY 993 KTLFLQEARLVKQWQFDDDLSTV 1015
DB 1352 rlprrllrnmiswlladdaiall 1374
RESULT 10
ID AAY14080
XX AAY14080 standard; Protein; 1295 AA.
AC AAY14080;
XX
XX 20-JUL-1999 (first entry)
DT
XX Gonadotropic hormone protein sequence.
DE
XX Gonadotropic hormone; excessive ovulation animal; transgenic animal;
KW

KW totipotent cell; somatic cell chromosome.
 XX Homo sapiens.
 OS JP11113444-A.
 XX
 XX PD 27-APR-1999.
 XX
 XX PF 14-OCT-1997; 97JP-0280830.
 XX
 XX PR 14-OCT-1997; 97JP-0280830.
 XX
 XX PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 XX (SAGA/) SAKAI H.
 XX WPI: 1999-320709/27.
 XX DR N-PSDB; AAX58001.
 XX
 XX PT An excessive ovulation animal - useful for improving the
 XX productivity of animals
 XX
 XX PS Claim 3; Page 14-18; 18pp; Japanese.
 CC This sequence represents a gonadotrophic hormone.
 CC The invention relates to an excessive ovulation animal, which is a
 CC transgenic animal with a totipotent cell containing a DNA fragment
 CC containing a promoter sequence and a gonadotrophic hormone coding
 CC sequence. The DNA fragment is in the somatic cell chromosome. The
 CC excessive ovulation animal is useful for improving the productivity of
 CC animals. The method can improve the productivity of a useful animal.
 XX
 XX SQ Sequence 1295 AA;
 Query Match 11.2%; Score 591.5; DB 20; Length 1295;
 Best Local Similarity 24.4%; Pred. No. 1.3e-40;
 Matches 237; Conservative 167; Mismatches 346; Indels 223; Gaps 38;
 QY 69 FLKSLKEMNYPLFQDLNGQ---SLFHQTSGLDLD----- 100
 DB 344 flgmks-saevrpdqsgelcelletsesnledsavgplvemaqgeaqvfgakn 402
 QY 101 LAODLKLITPSFLNF-----YPLGSDIDITNKSTFEPVLMKRDQHNHR 148
 DB 403 lneqlrayasafrrmsliddsdladhlqcdlsi---askhlskpv----- 449
 QY 149 VEOITLNGLOALOSPCTIEGSGKSTLQRIAMLMGSGKALKKFFVFFRLS-- 206
 DB 450 qeplvprevglnsvmcvegeasgkvtllkkaflwaagccpllnrfqvlvylsist 509
 QY 207 RAOGLEFETLCDOLLDPGTIRKOTFMAMLLKROVLFLLDGYNEF--KPONCPETIAL 264
 DB 510 rpdqglasiicdqllekgesvtemcmrnllqkngvllfllddykelcsjpr--vlgkl 566
 QY 265 IKENHRKKNVITVTTECLRHNGALTAEVGDMTEDSAQALIREVLKELA--EGLL 322
 DB 567 iqknhlsrtcllavtrnrdtryletllkafpyltvcllxrlfshnmrlrkfkm 626
 QY 323 LOIOKSRCLRMKTPFVVITCAIQGSEPHSHQTOTLPHFYDLLQKNKHKGVA 382
 DB 627 vyfgkngslqkqktrpfaivaicahwfgypfdsvdvaavkysymelstlnk----- 680
 QY 383 ASDPFR-SLDHCGDLEAGVFNKFELODV--SSVNEVDLLTGLCKYTAORFKPKY 439
 DB 681 taellkatvsscgealalkffscfendddlaeaygedelmcimskftaqlrlpfy 740
 QY 440 KFFHKSPQETTAGRRSLTSHPEEVTGNGYLOKMWISDITSYSSILKRYTCSSV 499
 DB 741 rfpafqeflaqmrllellsdqfegndqlylhkqinspmatvsaynflnyv--ssl 798
 QY 500 EATRA---VMKHAAVYOHGCLGLSLAKRPLRMROESLOSXKN-----TTEQEI 545
 DB 799 pscakgpkvlshlhlvdn-----keslensenddylkhpqelslqmq 843

QY 546 LKAI-----NINSEVEGGINL-----YOESTSKSALSOEFAFROGKSLYINSNIPD 593
 DB 844 lrglwgqcpqayfsmvsehllvalaktaygnt-vaacpffvqlfgrtlltgalnlg 901
 QY 594 YLFD-----FFEHLPNC--ASALDFIKLDYGG--AMASWE 625
 DB 902 yfhdhpeslslrshfprngktsprahfsvletcdksqypltdqyasafemewe 961
 QY 626 -----KAEEDTGGIHMEAPET-IPRAVSLFPMNMKQEFRTLEVTL 666
 DB 962 rnlakednvksymdqgraspalslsgywklspxkylp-----cleavd 1006
 QY 667 RDESKLNKODITVIGKIFSSATSLRLQIKRCAGVAGSLSVLSTCK--NIYSLWEASPLT 725
 DB 1007 ndldvvgqdmlellmvtvssagrfelhnrgfieslraplelskavtksisxlets 1066
 QY 726 IEDERHITSTVNKTSIH---DLQNRPLPGGLTDSGNKNTKLMDNIRK----- 775
 DB 1067 aaeqellltpslslevsgqlsqdqlfpr--ldkflclkelvdegnlnvsvlpee 1124
 QY 776 -----NEDAIKLAEGKLNKKMCLFHL--THLSDIGEGMDYIVKSLSSP 818
 DB 1125 fmfhmeklllqisaeyopskvlkqnsplnhvhlkcnfstdgsimtmivs----- 1179
 QY 819 PC-DLEBIQLVSCCLSANAVKILQNLMLVKLSIDL--SENYLEKDGNEALHELIDRMN 876
 DB 1180 -cklleikfsdsff--qavpfva-slpntfislkllnleqgqfdeetsekfayllgsls 1235
 QY 877 VLEQLRALMLPMG 889
 DB 1236 nlee--lllptg 1245
 RESULT 11
 ID AAY09540
 ID AAY09540 standard; Protein; 1295 AA.
 XX
 XX AAY09540;
 XX
 XX DT 20-JUL-1999 (first entry)
 XX
 XX DE Human apoptosis inhibiting protein #2.
 XX
 XX KW Human: apoptosis inhibitory protein; apoptotic disease; diagnosis;
 XX spinal muscular atrophy.
 XX
 XX OS Homo sapiens.
 XX
 XX PN JP11116599-A.
 XX
 XX PD 27-APR-1999.
 XX
 XX PF 14-OCT-1997; 97JP-0280831.
 XX
 XX PR 14-OCT-1997; 97JP-0280831.
 XX
 XX PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 XX WPI: 1999-323531/27.
 XX DR N-PSDB; AAX56273.
 XX
 XX PT New apoptosis inhibitory protein - useful for determining mechanism
 XX of various apoptotic diseases e.g. human spinal muscular atrophy
 XX
 XX PS Claim 1; Page 8-11; 16pp; Japanese.
 CC The present sequence represents a human apoptosis inhibitory protein.
 CC The apoptosis inhibitory protein is useful for the elucidation of
 CC the mechanism of various apoptosis diseases such as human spinal
 CC muscular atrophy and the diagnosis, the prevention and the treatment
 CC of such diseases.

XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00663.
 PE
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 PS Claim 27; SEQ ID No 30571; 654pp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP;
 CC see A1131315-A157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 73 AA;

Query Match 7.0%; Score 372; DB 22; Length 73;
 Best Local Similarity 100.0%; Pred. No. 5.2e-24;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 944 LAGNRVSSDGMALFMGVFENIKOLVFPDFSTKEFLPPALVRKLSQVLSKTLFLOEARLY 1003
 DB 1 LAGNRVSSDGMALFMGVFENIKQLVIFDFSTKEFLPPALVRKLSQVLSKTLFLOEARLY 60
 QY 1004 GWCFFDDDDLSVIT 1016
 DB 61 gwgfdiddlsvit 73

RESULT 15
 AAM05439
 ID AAM05439 standard; Protein; 73 AA.
 XX
 AC AAM05439;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Peptide #4121 encoded by probe for measuring breast gene expression.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PE 29-JAN-2001; 2001WO-US00661.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-476286/51.
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 PS Claim 27; SEQ ID No 14179; 322pp; English.
 CC The present invention relates to novel single exon nucleic acid probes
 CC (see A1100010-A110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at fip.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 73 AA;

Query Match 7.0%; Score 372; DB 22; Length 73;
 Best Local Similarity 100.0%; Pred. No. 5.2e-24;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 944 LAGNRVSSDGMALFMGVFENIKOLVFPDFSTKEFLPPALVRKLSQVLSKTLFLOEARLY 1003
 DB 1 LAGNRVSSDGMALFMGVFENIKQLVIFDFSTKEFLPPALVRKLSQVLSKTLFLOEARLY 60
 QY 1004 GWCFFDDDDLSVIT 1016
 DB 61 gwgfdiddlsvit 73

Search completed: March 25, 2002, 10:56:18
 Job time: 78 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 10:56:20 ; Search time 14.93 Seconds
(without alignments)
2514.721 Million cell updates/sec

Title: US-09-697-089-2
Perfect score: 5281
Sequence: 1 MNFKDNRSLRILQRMGMTVI.....MQPDDDDLSVITGAFKLVTA 1024

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100055

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	665.5	12.6	1403	1	BIRF_MOUSE
2	657	12.4	1402	1	BIRG_MOUSE
3	653	12.4	1403	1	BIRI_HUMAN
4	637.5	12.1	1403	1	BIRE_MOUSE
5	637.5	12.0	1447	1	BIRB_MOUSE
6	631.5	12.0	1403	1	BIRA_MOUSE
7	180.5	3.4	1078	1	C2TA_MOUSE
8	176	3.3	862	1	AACT_DICRI
9	168	3.2	3911	1	AKA9_HUMAN
10	164	3.1	1060	1	EG51_XENLA
11	152.5	2.9	2418	1	SPCA_HUMAN
12	150	2.8	1277	1	YMW6_YEAST
13	149.5	2.8	3210	1	CENF_HUMAN
14	148	2.8	2167	1	BEM2_YEAST
15	146.5	2.8	2493	1	YDA4_YEAST
16	146	2.8	1682	1	MSPI_PLA3
17	145.5	2.8	4196	1	DYHC_SCHRO
18	143.5	2.7	1315	1	CHAO_DROME
19	143.5	2.7	2869	1	RBP1_PLAIV
20	143	2.7	875	1	ZIP1_YEAST
21	141.5	2.7	1130	1	C2TA_HUMAN
22	141	2.7	2477	1	SPCN_CHICK
23	140.5	2.7	326	1	YVZ9_CAEEL
24	140	2.7	1875	1	MUPL_YEAST
25	139.5	2.6	567	1	YE28_MERVA
26	139.5	2.6	5430	1	ACF7_HUMAN
27	138.5	2.6	1526	1	MYS2_SCHRO
28	138.5	2.6	4540	1	DYHC_PARTE
29	138	2.6	2710	1	TOXA_CLODI
30	138	2.6	5327	1	ACF7_MOUSE
31	137	2.6	1256	1	FULI_DROME
32	136	2.6	999	1	RUK5_ARATH
33	136	2.6	1048	1	HPM1_YEAST

34	135.5	2.6	3321	1	KEND_HUMAN	095613 homo sapien
35	135	2.6	2230	1	GOG4_HUMAN	013439 homo sapien
36	134.5	2.5	1928	1	MYS1_YEAST	P08964 saccharomyc
37	134	2.5	567	1	GPV_RAT	O08770 rattus norv
38	134	2.5	959	1	VDP_RAT	P41542 rattus norv
39	133.5	2.5	456	1	RINI_RAT	P29315 rattus norv
40	133.5	2.5	1818	1	Z294_HUMAN	O94822 homo sapien
41	133	2.5	1325	1	G160_MOUSE	P55937 mus musculu
42	133	2.5	2238	1	RRLP_BUNTW	P20470 bunyamwera
43	132.5	2.5	582	1	SHO2_MOUSE	O88520 mus musculu
44	132.5	2.5	1057	1	EG5_HUMAN	P52732 homo sapien
45	132.5	2.5	1648	1	Y39H_YEAST	P47171 saccharomyc

ALIGNMENTS

```

RESULT 1
ID BIRF_MOUSE STANDARD: PRT: 1403 AA.
AC Q9JTB6: P81704; 009122; 009121;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1F (NEURONAL APOPTOSIS
DE INHIBITORY PROTEIN 6)
CN BIRCLF OR NAIP6 OR NAIP-RS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20414747; PubMed=10958627;
RA Endrizzi M.G., Hadinoto V., Growney J.D., Miller W., Dietrich W.F.;
RT "Genomic sequence analysis of the mouse Naip gene array.";
RL Genome Res. 10:1095-1102(2000).
RN [2]
RP SEQUENCE OF 82-168 FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=97131520; PubMed=8975718;
RA Scharf J.M., Damron D., Frisella A., Bruno S., Beggs A.H.,
RT Kunkel L.M., Dietrich W.F.;
RL "The mouse region syntenic for human spinal muscular atrophy lies
RT within the 1q91 critical interval and contains multiple copies of Naip
RT exon 5.";
RL Genomics 38:405-417(1996).
CC -1- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
CC SIGNALS.
CC -1- SIMILARITY: CONTAINS 3 BIR REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF242431; AAC82751.1; -
CC EMBL; U66327; AAC52975.1; -
CC MGD; MGI:1298222; Birc1f.
CC InterPro: IPR001370; BIR.
CC Pfam; PF00653; BIR; 3.
CC SMART; SM00238; BIR; 3.
CC PROSITE; PS01282; BIR_REPEAT_1; 2.
CC PROSITE; PS0143; BIR_REPEAT_2; 3.
CC Apoptosis; Repeat; Multigene family.
FT REPEAT 60 127 BIR 1.
FT REPEAT 159 227 BIR 2.
FT REPEAT 278 345 BIR 3.
SQ SEQUENCE 1403 AA; 159823 MW; 9DA912503358C4E9 CRC64;

```

Query Match	12.6%;	Score 665.5;	DB 1;	Length 1403;
Best Local Similarity	24.1%;	Pred. No. 1.3e-30;		
Matches 250;	Conservative 186;	Mismatches 414;	Indels 187;	Gaps 37;

0Y	101	LAODKJLUTHPSPUNF-----YPCEDDIDJFNKSPTEBPVLMRKQDHNHR	148
Db	403	LSBDRJTYKRTSPCHMNLPEVSSLSJCTDHLLOCYSII---SKIVSPV-----	449
0Y	149	VEOLJTLNGLLOALOSPICIEGESGKSTPLQRIAMJWSGCKALTKFKVEFLRIS-	206
Db	450	OGALTFIEPVNSLISWVCWGEAGSGKTFELKRIAFIAMSOGCCPLRYFOFLVYLSSI	509
0Y	207	RAQSGFETLCDDJLDIPGIRKQTFMAMJLTKRQVRJFLDQVNEFK--PONCEPEAL	264
Db	510	TPDQGLNITICTOLLGAGGCSIEVCLSSSTIQOJHOVFLFLDDYSGLASLPQ--ALHTL	566
0Y	265	IKENHRKNNVITYTTTECLRHIREGALTAIEYGDWTEDSAQALREVILKEI--AEGIL	322
Db	567	ITKNVLERJCTCLIAVHNRRJDRJPLGSLDLOEPFPIVNYFVLRKFFSHDIIQVELI	626
0Y	323	LOJQKSRCLNMLKTPLEVITC-AIQMBSEPHSTQTLTFHNYEDLLQKNKHKGV	381
Db	627	IYFSEKNDLOGVYKPLFVAAYCNDMNONASADDFQDVTTFHSMOYLSLKYK----A	681
0Y	382	AASFIRSLDHGGBLAEVESHKRPDELDV--SSVNDVJLTJGLLCKXTAORFKRY	439
Db	682	TAESLOQTVSSCGOALITGLTSSCFEPNSDDIAEKAGVDEDAVLTJFLMSKFTAOQLRPVY	741
0Y	440	KFHKSPOEYTAGRRJLSLJTSHEPEVYTKGNGYLOKMWISIDITVYSSILRYTCG--SS	498
Db	742	RFLGRLPQEOFLAARJLELLSDROQDQJGLYLRQIDSPKAIANSFNIFLYVSSHSS	801
0Y	499	VEATRAVAKHLAAYQHGCLLGISIAKRPJMRQESIQSKYKTTEDQ-----	544
Db	802	SKAAPTYVSHL-----LOJVDKESJENNSEMEDVYKJLHPQFLWFQ	843
0Y	545	-----IJKAININSFECCIHLYQ-----ESTSKSLSQEPFAFGKSYINSGNI	591
Db	844	FVRLQMLVSPESPSPV--SEHLRLALJFNAESNTVACSPFIQJLGRJLALRYNL	901
0Y	592	PQVLFDFFEHLPNCASALDFIKLDFYGGAMASWEKAEDTGGIHMEAPETYIDSRAYSL	651
Db	902	-EY---FMDH-PEJLILRLSKVISINGNMKSSYVDSEKTY--YFENLOPAINEEYISA	953
0Y	652	F---FNNKOER-----TLEVTJIDFESLN	673
Db	954	FEHVSEMRNRNPAODEETIKKYEINIMPRALPDISEGYWNLSPKPKCIKPEJQVANNMGPAD	1013
0Y	674	KQDITYJGKIFSSATSLRLOIKRCAGVAGSLSYLSTCK-NIYSJLMEVASPJTIEDERHI	732
Db	1014	QALLQVJMEVFSASQSTIEPHJFNSSQFELISIRALELSASVYTKCSMRNLSLSRAQJLL	1073
0Y	733	TSVTNKLTLST-----HDLON-----ORLPG-----LTDLSGNKJNLTK	767
Db	1074	LJTLPALQSLSEVENOLPDQLFHNHKKFLGJLKEJLQVRLDGRKPDVJLVPBEERJLNHMEK	1133
0Y	768	LIMNIMKMNEDDAIKLAEGLKJLKKKCLPHJNLSDIGSGMUYIKSLSSEPC-DLEJO	826
Db	1134	LSIT--STESDLSKLKFIQJPNJLVHJLK--CDFLSNCEJSLTALAS--CKKLREI	1187
0Y	827	LVSCLLSA-NAVJILAQNLINVLKLSJLDL-SENYLEKGNALHELIDRMNVLEJOJLAL	884
Db	1188	FSGQCFEAMFVNJLPR---NPFVSLKJLISLKGQGFADKRTSKFPAQJALGJSURJNE--L	1240
0Y	885	MLPWCQDVQGSLSLKLHLEVPOLYVJLGNKMRJLTDTEIRJLGAFEGKJNPLKJNFQJNL	944
Db	1241	LVPJGDGJHOVAKJLIVQCIQJLPCJLFLVAFHD-JLDDSEVIEIGEAATSGSFQJLENDI	1299
0Y	945	AGN-RVSSDGLJAMGVENJOKOLVFFDSTKJELP-----DPALVJKLSQJLSTJFLQ	998
Db	1300	SMNHKITEBGRNFQALDMLPJNLOJMLNICRN--JPGRIQOVQATTYVJALGHCVSRJPSLT	1357

```
QY 999 EARLVGWFDDDL SVI 1015
    :: | |:: | |
Db 1358 RLGMLSWLDEEDMKVI 1374
```

	RESULT	2	
BIRG_MOUSE	ID	BIRG_MOUSE	STANDARD; PRT: 1402 AA.
AC	O9JIB3;		
DT	20-AUG-2001	(Rel. 40, Created)	
DT	20-AUG-2001	(Rel. 40, Last sequence update)	
DT	20-AUG-2001	(Rel. 40, Last annotation update)	
DE	BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1G (NEURONAL APOPTOSIS INHIBITOR PROTEIN 7).		
CN	BIRCIG OR NAIP7.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid:10090;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	EMBL:U0614747; PubMed:10958627;		
RA	Emrizzi M.G., Hadnoto V., Gromey J.D., Miller W., Dietrich W.F.:		
RT	"Genomic sequence analysis of the mouse Naip gene array."		
RL	Genome Res. 10:1093-1102(2000).		
CC	-I FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF SIGNALS.		
CC	-I SIMILARITY: CONTAINS 3 BIR REPEATS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).		
CC	-----		
DR	EMBL: AF242433; AAF82749.1; -		
DR	MGI: MGI1858256; Birc1g.		
DR	InterPro: IPR001370; BIR.		
DR	Pfam: PF00653; BIR; 3.		
DR	SMART; SMO0238; BIR; 3.		
DR	PROSITE: PS01282; BIR_REPEAT_1; 2.		
DR	PROSITE: PS0143; BIR_REPEAT_2; 3.		
KW	Apoptosis; Repeat; Multigene family.		
FT	REPEAT 60 127 BIR 1.		
FT	REPEAT 159 227 BIR 2.		
FT	REPEAT 278 345 BIR 3.		
SO	SEQUENCE 1402 AA; 159662 MW; CIDFFBA359893E0D CRC64;		

Db 627 IYFSENKDLQVYKTPLEVAANCNDNMNANASQDDVTLFHSYQVLSLKYK-----A 681
 QY 382 AASDFIRSLDHGDLALBESVFSHKFDFELQDY--SSVNEVDLLTTGLCKYTAQREPKY 439
 Db 682 TMSLSQATVSSGGLATGLFSSCFEFNDDDLAEGVDDVAKLTTFIMSKFQAQNLRPY 741
 QY 440 KFFHKSFOEYTAGRRLSLTSHPEPEYTKNGYLOKNAVYSIDITSYSSLLRYTCG--SS 498
 Db 742 RFLGFLFOEFLAARLTLLSSDROEDDGLYLRLQIDSPKAINSFNFIYVSSHSS 801
 QY 499 VQATRAVYKHLAAVYQHGLGLSLAKRPLMRQESLQSKNTDE----- 544
 Db 802 SKAAPTVVSHL-----LQVDEKESLENNSEMEDYMKLHPQFLMFQ 843
 QY 545 -----ILKAININSFVECGIHLVQ-----ESTKSALSQEPFAFGKSLYNSGNI 591
 Db 844 FVRGLMIVSPESFSFV--SEHLRLALIFAESNTVAECSPFIQFLGRGLARLAVLT 901
 QY 592 PDYLPDFEHLDPNCASALDFIKLDYFGAMASWEKAEDTGIMHEAPETIYPSRAVSL 651
 Db 902 -EY---FMDH--PESILLRLSLKVSINGNMSSYVDYSFKT---YFENLQPPAINDEYTS 953
 QY 652 F---FNMKQEFR-----TLEVTLRDPSKLN 673
 Db 954 FEHVSEMRNRFQADEIITKNENIMPRALPDISEGYMNLSPKPKIPKLEQVNNMGPAD 1013
 QY 674 KODITYLGKIFSSATSLRLQIKRCAGVAGSLVSTCK-NIYSIMVEASPLTIDERHI 732
 Db 1014 QALQVLMVEFASQSIIEFHLENSGFLESTRPALELSKASTKCSMRLELSRAQELL 1073
 QY 733 TSVNMLKLTST-----HDLYN-----QRLPGG-----LTDLSGINKNLTK 767
 Db 1074 LTLPALQSLSEYSETNQLPDQLEFHNHKEFLGELKELCVRLDGKDYVLSYPEEFLNLHMEK 1133
 QY 768 LIMDNINKNNEEPAIKLAGSKNKKMKCLEFHLHSDIGBMGYIKYKSSSEFC-DLEEQ 826
 Db 1134 LSIRF--STESDLSKLVFTQMFNPLVHFHLK--CDPLSNCSLMTALAS--CKLREIE 1187
 QY 827 LVSCSLA-NAVKILIAONLHNLKLSILDL--SENYLERKDNALHELIDRMNVLEDTAL 884
 Db 1188 FSGQGFEMATFNILP-----NFVSLKILSLKQGFADKETSFKFQALGSLRNLE---L 1240
 QY 885 MLPWCCDVQGSLSLLKHLVEYPLVKLGKLNKWLDTDEIRILGAFPGKNPLKNFOOL-N 943
 Db 1241 LVPETDGHQVAKLIVROCLQPCRLVLAFAHILDESEYIEGATSG----SFQKLEN 1295
 QY 944 L---AGNRVSSGWLAFMGVFNELKQLVPEFDSKTEFLP-----DPALVRLKSLQVLSKLT 995
 Db 1296 LDISNMHKTTEGYNRFQALDNLPLQMLNLCRN--IPGRIQVQATTVKALCHCVSRLP 1353
 QY 996 FLOEARLVGWFDDDLVSI 1015
 Db 1354 SLTRGLGMLLDEEDMKVI 1373

RA Roy N., Mahadevan M.S., McLean M., Shuler G., Yaraagi Z.,
 RA Farahani R., Baird S., Besner-Johnston A., Lefebvre C., Kang X.,
 RA Saitoh M., Aubry H., Tamai K., Guan X., Ioannou P., Crawford T.O.,
 RA de Jong P.J., Surh L., Ikeda J., Korneluk R.G., Mackenzie A.,
 RT "The gene for neuronal apoptosis inhibitor protein is partially
 RT deleted in individuals with spinal muscular atrophy."
 RL Cell 80:167-178(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND REVISIONS.
 RC TISSUE-Brain.
 RX MEDLINE-98163755; PubMed-9503025;
 RA Chen Q., Baird S.D., Mahadevan M., Besner-Johnston A., Farahani R.,
 RA Xuan J.-Y., Kang X., Lefebvre C., Ikeda J.-E., Korneluk R.G.,
 RA Mackenzie A.E.;
 RT "Sequence of a 131-kb region of 5q13.1 containing the spinal muscular
 RT atrophy candidate genes SMN and NAIP."
 RL Genomics 48:121-127(1998).
 RN [3]
 RP SEQUENCE OF 386-623 FROM N.A.
 RA der Steege G., Draalgers T.G., Grootscholten P.M., Oelnga J.,
 RA Anzevino R., Vellona I., Brahe C., Scheffer H., van Ommen G.J.B.,
 RA Buys C.H.C.M.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 222-1403 FROM N.A.
 RA Jones K., Graves T., McPherson J.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP FUNCTION.
 RC TISSUE-Liver.
 RX MEDLINE-96149249; PubMed-8552191;
 RA Liston P., Roy N., Tamai K., Lefebvre C., Baird S., Cherton-Horvat G.,
 RA Farahani R., McLean M., Ikeda J., Mackenzie A., Korneluk R.G.;
 RT "Suppression of apoptosis in mammalian cells by NAIP and a related
 RT family of IAP genes."
 RL Nature 379:349-353(1996).
 CC -I- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
 CC SIGNALS.
 CC -I- TISSUE SPECIFICITY: EXPRESSED IN MOTOR NEURONS, BUT NOT IN SENSORY
 CC NEURONS. FOUND IN LIVER AND PLACENTA, AND IN A LESSER EXTENT IN
 CC SPINAL CORD.
 CC -I- DISEASE: MUTATED OR DELETED FORMS OF NAIP HAVE BEEN FOUND IN
 CC INDIVIDUALS WITH SPINAL MUSCULAR ATROPHY TYPE I (SMA TYPE I). SMAS
 CC ARE FATAL AUTOSOMAL RECESSIVE DISORDERS SUBCLASSIFIED AS TYPE I
 CC (MERDING-HOFFMANN DISEASE), TYPE II (INTERMEDIATE FORM), AND TYPE
 CC III (WOLFFHART-KUGELBERG-MELANDER DISEASE) BASED UPON THE AGE OF
 CC ONSET AND CLINICAL SEVERITY. THESE NEURODEGENERATIVE DISORDERS ARE
 CC CHARACTERIZED BY DEGENERATION OF LOWER MOTOR NEURONS, LEADING TO
 CC PROGRESSIVE PARALYSIS MUSCULAR ATROPHY. CONCERNS I IN 6000
 CC NEBORN.
 CC -I- SIMILARITY: CONTAINS 3 BIR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U19251; AAC52045.1; -
 CC EMBL: U80017; AAC52047.1; -
 CC EMBL: U21913; AAA64504.1; -
 CC EMBL: AC005031; AAC62261.1; -
 CC MIM: 600355; -
 CC InterPro: IPR001370; BIR.
 CC Pfam: PF00653; BIR; 3.
 CC SMART: SM00238; BIR; 3.
 CC PROSITE: PS01282; BIR_REPEAT_1; 3.
 CC PROSITE: PS0143; BIR_REPEAT_2; 3.
 CC Apoptosis; Repeat.
 FT REPEAT 60 127 BIR 1.
 FT REPEAT 159 227 BIR 2.

```

FT REPEAT 278 345 BIR 3.
FT CONFLICT 222 223 PK -> YR (IN REF. 4).
FT CONFLICT 386 387 VP -> ST (IN REF. 3).
FT CONFLICT 535 535 M -> V (IN REF. 3).
FT CONFLICT 553 553 Y -> H (IN REF. 3).
FT CONFLICT 1228 1231 MISSING (IN REF. 4).
SQ SEQUENCE 1403 AA; 159613 MW; 566304C154DA5E64 CRC64;

Query Match 12.4%: Score 653; DB 1; Length 1403;
Best Local Similarity 23.7%: Pred. No. 6.9e-30;
Matches 263; Conservative 197; Mismatches 410; Indels 238; Gaps 42;

QY 69 FLKSLKEMNPLFODLNGQ---SLFHQTSGLDD----- 100
DB 344 FLQWMS-SAEVPPDLQSRGLCELLETSESNEIDSAVPIVEMAQGBAOWQEAKN 402
QY 101 LAODKULYTPFLNF-----YPLGDDIDITNKSTFEPEVLMKDHHR 148
DB 403 LNEQLRAAYTSASFRHMSLDDISSDLATDHLGCDLSI---ASKHISKPV----- 449
QY 149 VEOTLNGLLAOLSPCITIGESGKSTLLORIAMLWSGCKALKPKFVFLRLS-- 206
DB 450 QEPVLVEVGNLNSVNCVEGASGCTVLLKTAFLMASGCCPLNRQVLVFLSST 509
QY 207 RAOGFLETLODLDIPGTIRKQTFAMMLKRLQVFLLDGYNE--KRONCPRIEAL 264
DB 510 RPDGLASIIICDQLLEKESVTECMENIIQQLKNQVFLLDVKEICSTPQ--VIGKL 566
QY 265 IKENHRKKNVYITTECLRHIOFAGALRAEVDGMDSDSAQALIREVLKELA--EGLL 322
DB 567 IOKNHLSTCLLAVNRNRRADIRYLETLEKAFPPYNTVCLRLKFSHNMTRLRKFM 626
QY 323 LQIQSRCLRMIMKTPFVVVITCAIOMGESEFHSHTQTLTFHTFYDLLIOKNKHGVA 382
DB 627 VYEKNGSLQKIQKTPFVAICAHMFQYPRDPFDVAVKSMELSLRNK-----A 680
QY 383 ASDPFR-SLHCGDLAEGVFSHKFDELQDV--SSYNEVDLLTTGLCKYTAORFKPY 439
DB 681 TAEIKATVSSCGELAKGFECSCEFFENDDLAEGVDEDLTMCMSKFTAORLPFY 740
QY 440 KFFKRSOETAGRRSLSLTSHPEEVTNGVYLOKAVSISDITSYSSSLRTGSSV 499
DB 741 RFLSPAQOEFLAGKRLLELSDROEIODGLYHLKQNSPMATVSAVNNFLNTV--SSL 798
QY 500 EATRA---VMKHLAAYVOHCGCLGLSLAKRPLMROESLOSAYKN-----TTEOEI 545
DB 799 PSTKAGPKIVSHLHLVDN-----KESLENISENDVYLKHQPEISLOMQL 843
QY 546 LKAI-----NINSVECGIHL-----YQESTKSALSOEFERFQOKSLYINSQNP 593
DB 844 LRGIMQICPOAYFSMVSEHLVLAALKTAQYOSNT-VAACSPFVLOFLOGRTLLGALND-Q 901
QY 594 YLFP-----FPEHLFNC--ASALDFIKIDFEGG--AMASWE 625
DB 902 YEFDPHPSLSLHSHRPIPNKNTSPRAHFSVLETCEDKSOVPTIDODYSAPFPMMEWE 961
QY 626 -----KAEDTGIIHMEAPETY--IPSRVSLFFNMKOEFTLEVTLL 666
DB 962 RNLAEREDNVKSYMOMQRASPDLSGTGWRKSPQYKIP-----CLEVDY 1006
QY 667 RDSKLLKODITYIGKFFSATSLRLQIKRCAGVAGSLVLSLCK--NIYSLWEASPLT 725
DB 1007 NDIIVGQDMLEITMTVFSASQRIELHNSRGFIESIRPALELSKASVTKCSISKLELS 1066
QY 726 IEDERHITSVNTKLSH---DIQONRLRGGLDSIGNKLNKLMDNKKM----- 775
DB 1067 AAEDELLTFLPSLESLSGTSQSDQIPV--LDKFLCKELSVLDGNNWFSVPIEE 1124
QY 776 -----NEDEAKLAEGKLNKKMLFHL--THLSDIGENDYIYKSLSE 818
DB 1125 FPNFHHMEKLLIQTSAEVDPSKLVKLLQNSNHLVPHLKCNFFSDFPSLMMVLVS----- 1179

```

```

QY 819 PC-DLEEIQLVSCCLSANAVKILAOQLHNLVKLSIIDL-SENYLEKDGNEALHELIDRN 876
DB 1180 -CQKLEIKFSDSEFF--QAVPFVA-SLPNFIISLKIINLEEOQFPDEBETSEKAVYIIGLSL 1235
QY 877 VLQKLAELMPCWCDQOGLSSLSLKHLEEVLPOLVKLGKMMRLTDR-----EIRIIGAF 931
DB 1236 NLEE--LILPTDGDYIRVAKLLIQQCQOLHCLRVLSFEFTLNDSDVETAKVAISGGF- 1291
QY 932 GKNPLKNFOQLNLAGN-RVSSDGMFLAFMGVFNILKQIVPEDFS---TKRELPDPALVRKL 987
DB 1292 -----QKLENLKLKSLINKITEBEGYRNFQALDMPNQLDLSRHRTTECKAKATYVKS 1346
QY 988 SQVLSKLTFLQEARLVGQWQFDDDLVSI 1015
DB 1347 SOCVLRILRILRLNMLSLDADIALL 1374

RESULT 4
BIRE_MOUSE
ID BIRE_MOUSE STANDARD; PRT; 1403 AA.
AC Q9R016; Q9R029; P81703; 009122; 009121;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN IE (NEURONAL APOPTOSIS
INHIBITORY PROTEIN 5).
GN BIRCIE OR NAIP5 OR NAIP-RS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxId=10090;
RX MEDLINE=99431676; PubMed=10501978;
RA Hwang S., Scharf J.M., Gromney J.D., Endritz M.G., Dietrich W.F.;
RT "The mouse Naip gene cluster on Chromosome 13 encodes several distinct
functional transcripts."
RL Mamm. Genome 10:1032-1035(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=99417674; PubMed=10486205;
RA Endritz M., Huang S., Scharf J.M., Kelter A.R., Wirth B.,
Kunkel L.M., Miller W., Dietrich W.F.;
RT "Comparative sequence analysis of the mouse and human Lgn1/SMA
interval."
RL Genomics 60:137-151(1999).
RN [3]
RP SEQUENCE OF 82-168 FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97131520; PubMed=8975718;
RA Scharf J.M., Damron D., Frisella A., Bruno S., Beggs A.H.,
Kunkel L.M., Dietrich W.F.;
RT "The mouse region syntenic for human spinal muscular atrophy lies
within the Lgn1 critical interval and contains multiple copies of Naip
exon 5."
RL Genomics 38:405-417(1996).
RC Lgn1
RT FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
SIGNALS.
CC -1- SIMILARITY: CONTAINS 3 BIR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF135492; AAD56764.1; -
CC DR EMBL; AF131205; AAD56760.1; -
CC EMBL; U66326; AAC52974.1; -
CC MGD; MGI:1298220; Bircie.

```


DR InterPro: IPR001370; BIR.
 DR Pfam: PF00653; BIR; 3.
 DR SMART: SM00238; BIR; 3.
 DR PROSITE: PS01282; BIR_REPEAT_1; 2.
 DR PROSITE: PS01483; BIR_REPEAT_2; 3.
 DR Apoptosis; Repeat; Multigene family.
 FT REPEAT 60 127 BIR 1.
 FT REPEAT 159 227 BIR 2.
 FT REPEAT 278 345 BIR 3.
 FT REPEAT 92 92 K -> R (IN REF. 1).
 FT CONFLICT 144 144 S -> R (IN REF. 1).
 FT CONFLICT 242 242 S -> G (IN REF. 2).
 FT CONFLICT 472 472 T -> A (IN REF. 2).
 FT CONFLICT 516 516 A -> D (IN REF. 2).
 FT CONFLICT 521 521 A -> T (IN REF. 2).
 FT CONFLICT 533 533 V -> A (IN REF. 2).
 FT CONFLICT 538 538 S -> I (IN REF. 2).
 FT CONFLICT 1092 1092 E -> D (IN REF. 2).
 FT CONFLICT 1129 1129 H -> L (IN REF. 2).
 FT CONFLICT 1137 1137 R -> Q (IN REF. 2).
 FT CONFLICT 1242 1242 V -> I (IN REF. 2).
 FT CONFLICT 1276 1276 D -> N (IN REF. 2).
 SO SEQUENCE 1403 AA; 159695 MW; B27F645043BC642 CRC64;

Query Match 12.1%; Score 637.5; DB 1; Length 1403;
 Best Local Similarity 23.8%; Pred. No. 5.3e-29;
 Matches 248; Conservative 191; Mismatches 407; Indels 195; Gaps 38;

OY 101 LAQDKLHYHPSFLNF-----YPLGEDIDIIIFNLKSTFTEPVLMRKDOHHNR 148
 DB 403 LSEQLRDVNTKATFRHMNLPEVCSLSIGDHLSCDVSI--SKHISQPV----- 449
 OY 149 VEQLTLNGLALQSPCLIEGSGKSTLQRIAMLMGSKCAALTKFVFFRLS-- 206
 DB 450 GQALTIPEVSNLNMVCMCEGTSGKTFKRIAFVLASGCCPLLYRQVFLSLSSI 509
 OY 207 RAQGLFETLDQLDIPGTIRKQTFMMLTKLRORLVFLFDLGVNEFK--PONCEIEAL 264
 DB 510 TPDOGLANICQLAGAGCISEVCLSSIQQLQHOVFLFLDDYGLASLPQ--ALHRL 566
 OY 265 IKENHFRKNMVIITTTTCLRIHROFALTAEVDMTEDSAQALIREVLIKEL--AEGIL 322
 DB 567 ITKNLSFTCLLIHAVTKNRDRIILYGLTSLIEQEPFNVTSVRKFFSHDIICEKLI 626
 OY 323 IQIQSRCLRNLMKTRPLFVITCA--IQMGSEFHSHTQTLFHFHYDILLQKNHKKHG 380
 DB 627 IYFIDNKDIOGYKTPFLVAAVCTDIQNASAO--DKFQDVTLFQSYMOYLSLKYR---- 680
 OY 381 VASDFIRSLDHGDLLEGVSHKDFEELQV--SSYNEVDLLTGLLCKYTAQRFKPK 438
 DB 681 ATAERLQATVSSCGDLATGLFSSCFERNSDILAAGVDEDEKLTLLMSKTTAQRRLPV 740
 OY 439 YKFEHKSQOETAGRRSLSLTSHEPEVYKNGYLOKMWISIDITSYSSLLRYTCG-S 497
 DB 741 YRFLPLPFOEFLAAVRIETELSSDRQEDODGLYLRQIDSPKAIANSFNIPLEVYSSHS 800
 OY 498 SVEATRAVMKHLAAVYQCGCLLSIAKRPLMROESLOSVKNTTQELIKANINISFV-- 555
 DB 801 SSKAAPTVVSHL-----IQLVDEKESLEMSNENEYMKLHPOTFLWF 842
 OY 556 --ECGIHYOESTSKALSQE-----FEA-----FPOGSLYINSNP 592
 DB 843 QPVRGLIMLVSPSSSSSFVSEHLRLALIFAVESNVAACSPRILOFLNGKTLALVNLN- 901
 OY 593 DYLDFEFHLPNCASALDPFIKIDFYGGAMASWEKAEDTGTGIMEADPTVYIPSAVSLF 652
 DB 902 ---OYFRDHPESSLILSLKVSINGNMSSYVDYSFT--YFENLOPPAIDEYTSAF 954
 OY 653 ---FMWKEPRLLEVTLRDES-----KINKOD----- 676
 DB 955 EHISEWRKRFADDEETIKNYENIRPRALPDISEGYWKLSPKPKCIKPLEVOVNTDADQ 1014

OY 677 --ITYLGKIFSSATSRLQIKRCAGVAGSLVLTCK--NIYSLMVEASPLTFEDERNIT 733
 DB 1015 ALLQVLMVEFASQSQSTIEFFLPMSSGFLSEICPALSLKASVYKCSMSRLSLRAQDELL 1074
 OY 734 SVYTNKTYSTI-----HDIQN-----ORLPQG-----LTDISGLNKLNLTKL 768
 DB 1075 TLPALOSLEVSEFTNQLPEQLFHNLHKFLGKELCYRLDQKPDVLSVLPQEPFNHLMKRL 1134
 OY 769 IMDNKNMNEDEAIKLABGLKNLKKCLPHLTLSDIGSEMIVYVSLSEPC-DIEEIQ 827
 DB 1135 SIRT--STESDSKLVKFLQNPNNLHVFLK--CDPLNCSESIAMAVLAS--CKKRIEIEF 1188
 OY 828 VSCCLSA--NAVKILQNLNHLNVLKSLIDP--SENYLEKQGNVLAHLEIDPMNVLEQUTALM 885
 DB 1189 SRCREAFMFFVILP-----NFYSKILINKDQGFDPKETSSEFQAOLGSLRLEE---LL 1241
 OY 886 LPMGCDVQGSLSLLKHLHEEVPOLVKGILKNRMLTDEIRIL----GAFGKNPLKNQ 940
 DB 1242 VPTGDGIHQVAKLIVROCIQLPECLRLVLFPHDILDDSVIEIRAMTSGF-----QKLE 1295
 OY 941 QNLNLAGN--RVSSDGLAFMGVFNENKOLVFPDFSTKEFLP-----DPALYKRLSQVLSKL 994
 DB 1296 NLDISMNHRTIEGVRNFEQALDNLPLNQ--ELNICRNIPIGRIOVATTYKALGQCVSRL 1353
 OY 995 TFLQEARLVGMQFDDDLDSVI 1015
 DB 1354 PSLIRLHMISMLDDEDMKVI 1374

RESULT 5
 BIRB_MOUSE STANDARD: PRT: 1447 AA.

ID BIRB_MOUSE
 AC 090UK4: Q9R030; 009124;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1B (NEURONAL APOPTOSIS
 INHIBITORY PROTEIN 2).
 GN BIRB1B OR NAIP2 OR NAIP-RS6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99431676; PubMed=10501978;
 RA Huang S., Scharf J.M., Growney J.D., Endrizzi M.G., Dietrich W.F.;
 RT "The mouse Naip gene cluster on Chromosome 13 encodes several distinct
 functional transcripts.";
 RL Mamm. Genome 10:1032-1035(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99315342; PubMed=10384056;
 RA Yaraqhi Z., Diez E., Gros P., Mackenzie A.;
 RT "cDNA cloning and the 5'genomic organization of Naip2, a candidate
 gene for murine Legionella resistance.";
 RL Mamm. Genome 10:761-763(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=99417674; PubMed=10486205;
 RA Endrizzi M., Huang S., Scharf J.M., Kelter A.R., Wirth B.,
 RA Kunkel L.M., Miller W., Dietrich W.F.;
 RT "Comparative sequence analysis of the mouse and human Lgn1/SMA
 interval.";
 RL Genomics 60:137-151(1999).
 RN [4]
 RP SEQUENCE OF 82-168 FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=97131520; PubMed=8975718;
 RA Scharf J.M., Damron D., Fritsella A., Bruno S., Beggs A.H.,
 RA Kunkel L.M., Dietrich W.F.;
 RT "The mouse region syntenic for human spinal muscular atrophy lies

RT within the 1gnt critical interval and contains multiple copies of Nalp
 RT exon 5.
 RL Genomics 38:405-417(1996).
 CC -1- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
 CC SIGNALS.
 CC -1- SIMILARITY: CONTAINS 3 BIR REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC
 CC EMBL: AF135489; AAD56761.1; -
 CC EMBL: AF135490; AAD56762.1; -
 CC EMBL: AF102871; AAC73002.1; -
 CC EMBL: AF131205; AAD56759.1; -
 CC EMBL: U66329; AAC52977.1; -
 CC MGD: MGI:1298226; Birc1b.
 CC InterPro: IPR001370; BIR.
 CC Pfam: PF00653; BIR; 3
 CC SMART: SM00238; BIR; 3.
 CC PROSITE: PS01282; BIR_REPEAT_1; 2.
 CC PROSITE: PS0143; BIR_REPEAT_2; 3.
 CC K1 Apoptosis; Repeat; Multigene family.
 CC FT REPEAT 159 127 BIR 1.
 CC FT REPEAT 60 127 BIR 1.
 CC FT REPEAT 127 127 BIR 2.
 CC FT REPEAT 278 345 BIR 3.
 CC FT REPEAT 377 377 D -> G (IN REF. 3).
 CC FT CONFLICT 403 403 L -> F (IN REF. 3).
 CC FT CONFLICT 478 478 L -> I (IN REF. 3).
 CC FT CONFLICT 540 540 N -> Y (IN REF. 3).
 CC FT CONFLICT 862 862 K -> N (IN REF. 3).
 CC FT CONFLICT 1079 1080 SD -> FN (IN REF. 3).
 CC FT CONFLICT 1089 1089 R -> C (IN REF. 3).
 CC FT CONFLICT 1115 1115 K -> E (IN REF. 3).
 CC FT CONFLICT 1122 1122 T -> A (IN REF. 3).
 CC FT CONFLICT 1136 1136 D -> E (IN REF. 3).
 CC FT CONFLICT 1157 1157 S -> G (IN REF. 3).
 CC FT CONFLICT 1167 1167 G -> R (IN REF. 3).
 CC FT CONFLICT 1271 1271 F -> C (IN REF. 3).
 CC FT CONFLICT 1447 1447 AA: 164033 MW: 9566673BA60A2 CRC64;
 CC SEQUENCE
 Query Match 12.0% Score 632.5; DB 1; Length 1447;
 Best Local Similarity 24.2%; Pred. No. 1.1e-28;
 Matches 268; Conservative 195; Mismatches 423; Indels 221; Gaps 44;
 QY 38 VNICECEKVDADARGIIMILKKGSESCNLFKSLKEMNYPFLFDLNGSLFPHOTSGD 97
 QY 404 VSYLCNRDQ-DHSEAGQ-----RGCCASGTYLPS-----TDL-GGSEQWLOEA- 444
 QY 98 LDDLAODKLQLYHTPPLNF-----YPLGEDDIIFNLKSTFTEPYLMKRDQ 145
 QY 445 -RLSEQLRDTYKATFRHNNLPEVYSLGTDLHLSGVII--SKHISQPV----- 493
 QY 146 HHNEVDLTGLLQALOSPGIIEGSKGKSTLLQRIAMGSKKALKKFFVFLRL 205
 QY 494 ---GGSTLPIEVSNNLSVNCVEGASGKTPYLKRIAFIWAASCCPLNRFDLVFLSL 550
 QY 206 SRAAGG--LPEFLCDLDDIPGTIRKOTFAMMLKLKQRLVFLDGNERK--PONCEI 261
 QY 551 SSTPGEGLAKITICQALGAGGCISEVCLSSITIQDLQHVLLFDLDDYSGLASLPQ---AL 607
 QY 262 EALIKENHRKNNVIYTTTECLRNHRQFGALTAEVGDMTDSQAOLIREVLLKEIAEGL 321
 QY 608 HTLITKNYLSRTCLLIVHTNKKVGRIPYLDTSLEIKFEPFYMTVSVLRKLESHD----- 662
 QY 322 LLDIQR-----SRCRLNMLKTPPLFVYTICAIOMGSESESHQTLLFTFDLLIQKN 374
 QY 663 IMVRKFINFGFHEELQGIHKTPPLFVAACVCTDWFKNPSDQPPDVALFKAYMOYL----- 718

QY 375 KHKHGVASDFIRSLDHCODLLEGVSHKPFDELQV--SSVNEVDLLTGLCKYTA 432
 QY 719 SLKHKG-AAKPLQATVSSCGQALATGLFSSCFEPNSONLAEAGVDEDELLTCLMSKFTA 777
 QY 433 QREPKYKFFPKHSPOEYTAGRLSLTSHPEBEVTKGNGYLOKMWISIDITSTYSSLR 492
 QY 778 QRRLPVRFRGPPLOEFLAVRLTELLSSDROEDODGLTYLRQINSPLKAMSITYHTFLK 837
 QY 493 YTCG-SSVEATRAVMKHLAVYQHGLGLSIKRPIMRQESLOSAYKNTQOEILAINI 551
 QY 838 YVSHPSKKAAPVYVSHL-----LQYDEKESLENNSENNDYKILAP 879
 QY 552 NS--FVDCGITHLYQ-----ESTSKSALSQEPAPFGKSLYI 586
 QY 880 EALLMEICLGLMGWQSPESPLFISENLRLICLNFHESNTVAACSVYLIQFLRGRTLD 939
 QY 587 NSGNIPDYLFDFEBHENCASALDFIKLDYGGAMA-----SWEA----- 627
 QY 940 KVLST-QY--FMDH-PETLLLLKSIKISLNGNMWQRIOPISLEKSFKEKVOPTIDQY 994
 QY 628 -----AEDTGTHMEAEPEYIPRAVSLFENWK-----QEFRTLEVTLRD 668
 QY 995 AIAFQPINEYQKNLSEKKHITIKYEDMKHQPILNISTGY--WKLSKPYKIPPLEVQVTN 1052
 QY 669 FSKLNKODITYLGKIFSSATSLRLQIRKAGVAGSLVLTSCK-NIYSLMVEASPLTIE 727
 QY 1053 TGPADQALLQVIMEVFASQISIEPRLSDSGPLESIRPALSELKASVTKCSMRLELSRE 1112
 QY 728 DERHITSVTKLSTISHDLONRLPGGLTDSLGNLKNLTYLI----- 769
 QY 1113 DQKLLTLPTLOSLEVS--ETNQLPDLFHNHKLFTOLKEICLRYLDSKRPVLSVLPGEFP 1170
 QY 770 -----MDNIKM--NEEDATLAGLKNLKKMCLFHL--THLSDIGEMDYVSLSEPC 820
 QY 1171 NLHMEKLSIRSTESDLSLVKLIQNSPNLHVFLHLCNFLSC-DEPLTVLASCCK--- 1226
 QY 821 DLEIQLVSCCLSA-NAVKILIAONLNLVLTSLDL-SENYLKEGNEALHELIDRMNVL 878
 QY 1227 -LREIFSGRCFEMAFVNIPLP-----NFVPLKILNLRDQGFPPKETSERKPAVALGSLRNL 1281
 QY 879 EQLTALMLPMGCVQVQSSLSLKHLEVPQVYKLGKNNMLTTEIRIL-GAPFGKPLK 937
 QY 1282 EK--LFPPTGDDGHQYAKLIVRQCQLPCLRLVLAETLDDSDVLEIAGARG----- 1333
 QY 938 NFOOL-NL--AGNRVSSDGMFLAPMGVFENLKQLVFDFSTKEPLPD-----PALYKLS 988
 QY 1334 GFQKLEMLDLTLNKHITEEGRNPFQVLDNLPNLKNDIS--RHIEPCIOIQAITYKALG 1391
 QY 989 QVLSKLTFIQEARLVGMQFDDDLSTV 1015
 QY 1392 QCVSRPLSLTRLGLMSVLDDDEDIKVI 1418
 RESULT 6
 BIR_MOUSE
 ID BIR_MOUSE STANDARD; PRT: 1403 AA.
 AC 090WK5: 09R017: 09JIB5;
 DT 20-AUG-2001 (rel. 40, created)
 DT 20-AUG-2001 (rel. 40, last sequence update)
 DT 20-AUG-2001 (rel. 40, last annotation update)
 DE BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 1A (NEURONAL APOPTOSIS
 DE INHIBITORY PROTEIN 1).
 GN BIR3A OR NALP1 OR NALP.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Varaghi Z., Korneluk R.G., Mackenzie A.E.;
 RT "Cloning and characterization of the multiple copies of the murine
 homologue of Nalp (neuronal apoptosis inhibitory protein).";

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-99431676; PubMed-10501978;
 RA Huang S., Scharf J.M., Growney J.D., Endrizzi M.G., Dietrich W.F.,
 RT "The mouse Nalp gene cluster on Chromosome 13 encodes several distinct
 functional transcripts.";
 RL Mamm. Genome 10:1032-1035(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-20414747; PubMed-10958627;
 RA Endrizzi M.G., Hadinoto V., Growney J.D., Miller W., Dietrich W.F.,
 RT "Genomic sequence analysis of the mouse Nalp gene array.";
 RL Genome Res. 10:1095-1102(2000).
 CC -1- FUNCTION: PREVENTS MOTOR-NEURON APOPTOSIS INDUCED BY A VARIETY OF
 SIGNALS.
 CC -1- SIMILARITY: CONTAINS 3 BIR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF007769; AAB69223.1; -
 DR EMBL: AF135491; AAD56763.1; -
 DR EMBL: AF242432; AAF82752.1; -
 DR MGI: MGI:1296223; Birc1a.
 DR InterPro: IPR001370; BIR.
 DR Pfam: PF00653; BIR. 3.
 DR SMART: SM00238; BIR. 3.
 DR PROSITE: PS01282; BIR_REPEAT_1; 1.
 DR PROSITE: PS0143; BIR_REPEAT_2; 3.
 KW Apoptosis; Repeat; Multigene family.
 FT REPEAT 60 127 BIR 1.
 FT REPEAT 159 227 BIR 2.
 FT REPEAT 278 345 BIR 3.
 FT CONFLICT 343 343 I -> V (IN REF. 2).
 FT CONFLICT 359 359 E -> K (IN REF. 2).
 FT CONFLICT 624 624 D -> E (IN REF. 3).
 FT CONFLICT 1092 1092 D -> N (IN REF. 3).
 FT CONFLICT 1116 1116 G -> R (IN REF. 3).
 FT CONFLICT 1123 1123 L -> H (IN REF. 1).
 FT CONFLICT 1129 1129 T -> M (IN REF. 2).
 FT CONFLICT 1140 1140 A -> V (IN REF. 3).
 FT CONFLICT 1269 1269 A -> V (IN REF. 3).
 SQ SEQUENCE 1403 AA; 158692 MW; B31630259595EB67 CRC64;

Query Match 12.0%; Score 631.5; DB 1; Length 1403;
 Best Local Similarity 24.7%; Pred. No. 1.2e-28;
 Matches 264; Conservative 185; Mismatches 409; Indels 211; Gaps 45;

QY 93 TSEGLDLD-----LAODLKDYHTPSLFNF----- 117
 DB 371 TSESHDDPAAVHTVVGIGRSEAOQFOEARKSLSEQLNDYTKAFRIHNPBVCSSLCT 430
 QY 118 -VPLGEDIIDIFNLKSTFEPYLMKRDHHRVEQLTNGLLQALQSPCIIEGESGKRS 176
 DB 431 DHLICDVSII---SKHISQPV-----QGALTIPVSNLSWVCVGETSGKRT 477
 QY 177 TLLORIAMLMGSKKALKTKFVFFLRS--RAQGFETICDLDLPIGTRKOTEMA 234
 DB 478 TFLKRIATFLMAGCCPRLRFQVLYLSISTTPDGLANIICADLLAGGCSISEVCLS 537
 QY 235 MLTKRORVLEFLDGYNEPK--PQNCPEIALIKENHREKNNVITVTTTECLRHTRQGA 292
 DB 538 IIOQLQHOVFLFLDDYSGLASLPQ---ALHTLITKNYLSRTCLLLAVHTNRYKGRSTYD 594
 QY 293 LTAEVGDMTEDSAQALIREVL---IKELAEGLILQIOKSRCLRLNLMKTPPLFVVTATQIM 349

DB 595 TSLTEIKPEPLSNVYILKKFESHNIKRLLE-FMVYFGONEDLOGIHKTPLEVAAVCTDMF 653
 QY 350 GSEFESHSHQTLLPFTFDLLQKKHKKHGAADPFRSDHCDLALBEGVFSKRPPE 409
 DB 654 ENPSDQPPQDMALFKFSYMOYL---SLKHKG-AAAPLOATVYSSCQQLALGLFSSCFEEN 708
 QY 410 LQDV---SSVNEVDVLLTGLCKYTAORPKPKFKFHSKFOETAGRRSLSTSHPEEV 467
 DB 709 SDDLAEAGVDEDEBELTCLMSKFTQARLRPYRFLGFLPQEFELAMRLTELLSSRDQDQ 768
 QY 468 TKGNCYLOKMWISDITSTYSSSLRYTCG-SSVEATRAVMKHLAAYHGCIL----- 519
 DB 769 DLGLYLRQINSPALKATLYNNFLKYVSHPSKAGPTVSHLHLVDETELLENTYKNE 828
 QY 520 -----GLS-TAK--RPLMRESLSQYKNTTEQILKAIN-----NSFVEGHILYQ-- 563
 DB 829 DVVNHPPGTSRIKMLKELMLSPYSSFVSEHLRLALNFAVSNVAVCESPIILOFL 888
 QY 564 --ESTSKSLSQEF-----EAFQGSKL--YINSGNIP--DY--LPDFEHL-----P 603
 DB 889 RGRRLALAVLNQYRDRPDESILLVYSLEYSINGKVKYVDYVMEKSFETLPPTIDQ 948
 QY 604 NCASALDFETKIDFYGGAMASWEKAEDTGGIHMEBAPETTYIPSAVSLF----- 653
 DB 949 DYASAFE-----QKMEHEK-----NLSENEETI--SKIKNIFFLOPKISSG 988
 QY 654 NKKQ-----EFTLEVTLRDPSKLAKODITYLGKTFSSATSIRLDQKACVAGSLVYL 708
 DB 989 YKLSPPKPKCKIPRLLEVGYVNMGPADALLQVLMEFVASQSIEFRLSDSSGFLSIRPAL 1048
 QY 709 STCK-NIYSLWAEASPLTIEDERHITSVNLKTLTSHLDONORLPGLTDSL----- 759
 DB 1049 ELKSAVTKCSMRLELSRADELLITPALQSLVYS--ETQPLDQFLPHNHLKGLIKE 1106
 QY 760 -----GNLKNLTFLKIMDNIMK--NEEDAIKLAELKLNKMKCLFHLTR 800
 DB 1107 LCVRLDGKPDVLSVLPGEFPNL--LHMEKLSIRSTESDLSLVYLTIONSPRLHVEHLK- 1163
 QY 801 LSDIEBGNDIYKSLSSSEPC-DLEIQLVSCLSANA-VKIIAOLNHLVLSITDL-SE 857
 DB 1164 -CDPLSNCDSLMAVLAS--CKKLEIEFSGRCFEEMPVNILP---NFISIKTLNLSQ 1216
 QY 858 NYLEKDGNEALHELIDRNNVLEQLTALMLPMGCDVQGLSLSLKLHLEVPOLVGLKKNW 917
 DB 1217 QEPDKETSEKFRQALGSLRNEE---LLVPPDGGIHQYAKLIVRCLOLPLCLRYLAHFYI 1273
 QY 918 RLDTETIRIL-----GAFPGKNPLKRFQOLNLAGN-RVSSDGLWAFMGVFENLQOLVFE 971
 DB 1274 LDNDVIEIARVATVSGGF-----QKLEKLDLSNMHKITTEEGYRNFQALDNLPRILQ--N 1325
 QY 972 FSTKEFLPD-----PALVRKLSQVLSKTLFQEARLVQWQDDDDLSVI 1015
 DB 1326 LNICHIPECTIOVATYKALGQCVSRPLSLRHLMLSLMLDEEDMKVI 1374
 RESULT 7
 C27A_MOUSE STANDARD: PRT: 1078 AA.
 AC P79621; Q31115;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MHC CLASS II TRANSACTIVATOR C11TA.
 GN MHC2TA OR C11TA OR C27A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ND; TISSUE=Spleen;
 RX MEDLINE=97152412; PubMed=8995190;
 RA Sims T.N., Elliott J.F., Ramassar V., Denney D.W. Jr., Halloran P.F.;

RT "Mouse class II transactivator: cDNA sequence and amino acid
 RL comparison with the human class II transactivator."
 RN Immunogenetics 45:220-222(1997).
 [2]
 RP SEQUENCE OF 878-1020 FROM N.A.
 RC STRAIN-BALB/C; TISSUE-Thymus;
 RX MEDLINE-96188886; PubMed-8620527;
 RA Panelli M.C., Wang E., Shen S., Schluter S.F., Bernstein R.M.,
 RA Hersh E.M., Stoeck A., Gangavalli R., Barber J., Jolly D.,
 RA Akporiaye E.T.;
 RT "Interferon gamma (IFNgamma) gene transfer of an EMT6 tumor that is
 RT poorly responsive to IFNgamma stimulation: Increase in tumor
 RT immunogenicity is accompanied by induction of a mouse class II
 RT transactivator and class II MHC."
 RL Cancer Immunol. Immunother. 42:99-107(1996).
 CC -I- FUNCTION: ESSENTIAL FOR TRANSCRIPTIONAL ACTIVITY OF THE HLA CLASS
 CC II PROMOTER; ACTIVATION IS VIA THE PROXIMAL PROMOTER. NO DNA
 CC BINDING OF IN VITRO TRANSLATED CITRA WAS DETECTED. MAY ACT IN A
 CC COACTIVATOR-LIKE FASHION THROUGH PROTEIN-PROTEIN INTERACTIONS BY
 CC CONTACTING FACTORS BINDING TO THE PROXIMAL MHC CLASS II PROMOTER,
 CC TO ELEMENTS OF THE TRANSCRIPTION MACHINERY, OR BOTH. ALTERNATIVELY
 CC IT MAY ACTIVATE HLA CLASS II TRANSCRIPTION BY MODIFYING PROTEINS
 CC THAT BIND TO THE MHC CLASS II PROMOTER (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 060633; AAB48859.1; -
 DR EMBL; 046562; AAB05004.1; -
 DR MGI; 108445; C2ta.
 DR InterPro: IPR003592; LRR_out.
 DR SMART; SM00370; LRR; 4.
 KW Transcription regulation; Activator; ATP-binding.
 FT DOMAIN 52 138 ASP/GLU-RICH (ACIDIC).
 FT NP_BIND 368 375 ATP (POTENTIAL).
 FT CONFLICT 917 917 G -> R (IN REF. 2).
 SQ SEQUENCE 1078 AA; 118789 MW; 28FEA61BBF581378 CNC64;

Query Match 3.4%; Score 180.5; DB 1; Length 1078;
 Best local similarity 19.7%; Pred. No. 0.0062;
 Matches 176; Conservative 126; Mismatches 304; Indels 289; Gaps 42;

QY 151 QLTNLGLAQ--SPC-----IIEGSGKSTLLQRIAMGSGKCALTEFE 199
 DB 339 QLAHGLAEVLQVSDCRREGTQYVAVLGAKGSGKSHMARTVSHWA--CGQLQYDF 395
 200 VFELR--LSRAQG--GLFETLC-----DQLDIGTIRKQTFMALKLRQVL 244
 DB 396 VFYVPCGLDRPDYTLRLRLCPRLPSLOPLAMDEVID-----YIYQPRVL 443
 QY 245 FLIDGYNFRPQNCPELIEALIKENHFRKNVIVTTTECLRIHQFGALTAEVGMTED- 303
 DB 444 LILDAFEELDAQ-----GLLHGPCGSLSPER 470
 QY 304 -SQAQILRELIKELAEGL--LLLOQKSRCLNLMKTPLEVVITCAIQMGSEFHSHTQ 359
 DB 471 CSIRGLAGLFGKRLKGLTLLTPRGRSLAOSLSKAD-----AIFEVPSSTQAK 523
 QY 360 TTFHTFYDILIOKNKHKHGVASDFIRSLDHCGLALEG-----VFSHKPFDELQVS 414
 DB 524 TYMRHFEENGSTAGNOKALGL-----LEGQPLCGYSR- 557
 QY 415 SVNEDVLLTTGLCKYTAQFKPKYKFEHFKSFQETAYAGRISSLTSS-----HEPEEYFK 469
 DB 558 -----SPVVCRAVQC-----LSKALLEQGEADLPCTLTGLYSLGLCPAONSS 600
 QY 470 GNGYLOKMWIS-DITSTYSSLRTYTCGSSVEA-TRAVMHLAAYVQHGLLGLSTAKRP 527

DB 601 PCALVELAKLAEWELGRHSHOSTLQETRFSSVEYKTAWVQGL----- 642
 QY 528 LMRQESLOSAYKNTQEDELKAININSP-VEG---GIHLVQESNSK-SALSQEFAPFOGK 592
 DB 643 -----WQOTLETTAOLA-----FSSFLQCLGAWLACQNEIKKELEFOYALTPRKK 692
 QY 583 SLXIN-SGNIPDLDFE-FEHLNFCASAL-----DEFLKDFYGAMAS 623
 DB 693 RPDNMIEGPRLAGIVFPRAHCGALVEPAVNAVDKQKVLRYRLKRLGLTLRAG 752
 QY 624 -----WEKAEDTGCIHME-----EAPETIYPSRAVSLFFNMKOEFR 660
 DB 753 RLLELLHCAHETQPGIWEHVAHQLPQ-HLSFLGTRLTLPDYYVLGRA----- 799
 QY 661 TLEVTLRDESKLWKO---DITYLGKI--FSSATSLRQIKRCGVAGSL-----SIVLSF 710
 DB 800 -LETASODFSLDRQGVESFGLNVLGSLCVTSFRASLSDYMAWLSLQOQGEAQLQA 858
 QY 711 CKNIYSL--WEASPLTIEDERHITSVNLKITS-----IHLQNOQLPG- 754
 DB 859 AEEKFTIEPKAKSPKQVEDLDRLVQGRLRNPSSEDAKDLPAIRDLKLEFALGPLGP 918
 QY 755 -----LTDLSGNLKNFLKIMDNI---KNEEDAIKLAEGKMLKMLFHL--HLSDI 804
 DB 919 QAEPTLAKLIPAFSSLOHLDLDSLSENKIDKGVSKLSATFPOLKALETLMLSQNNITDV 978
 QY 805 G-----EGMDIYKSLSESPCEDEIQLVSCSANNVAKITLQNLHNLVLTSLDSENY 859
 DB 979 GACKLAELALAKSL-----RSLYNNCICDGAKSLAQLVDPMSLSLRMDVQFNK 1031
 QY 860 LEKQNEALHELIDRNVLEOLT--ALMLPWCQDVQGSLSLLKHUEVQVLKL 912
 DB 1032 FTAAGAOQLASSQKCPQVETLAMWPTPIPG-----VQEHQLQDLARISL 1077

RESULT 8
 ID AACT_DICDI STANDARD; PRT; 862 AA.
 AC P05095;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALPHA-ACTININ, NON-MUSCULAR (F-ACTIN CROSS LINKING PROTEIN).
 GN ABPA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxId=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX2;
 RX MEDLINE-87304850; PubMed-3622778;
 RA Noegel A., Witke W., Schleicher M.;
 RT "Calcium-sensitive non-muscle alpha-actinin contains EF-hand
 RT structures and highly conserved regions."
 RL FEBS Lett. 221:391-396(1987).
 RN [2]
 RP SEQUENCE OF 92-505 FROM N.A.
 RC STRAIN-AX2;
 RX MEDLINE-86304574; PubMed-3745276;
 RA Witke W., Schleicher M., Lottspeich F., Noegel A.;
 RT "Studies on the transcription, translation, and structure of alpha-
 RT actinin in Dictyostelium discoideum."
 RL J. Cell Biol. 103:969-975(1986).
 CC -I- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR
 CC ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING
 CC PROTEIN.
 CC -I- SUBUNIT: HOMODIMER, ANTIPARALLEL.
 CC -I- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
 CC -I- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY DOMAINS (CH).
 CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -I- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.

RT "Characterization of a novel giant scaffolding protein, CG-MAP, that
RT anchors multiple signalling enzymes to centrosome and the golgi
RT apparatus";
RT J. Biol. Chem. 274:17267-17274(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kemmer W.A., Deiss S., Schwarz U.;
RT "Cloning of Hyperion.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RA MEDLINE-99115654; PubMed-9915845;
RX Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RA Trotter K.W., Milgram S.L., Goldenring J.R.;
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
RT associated with centrosomes.";
RL J. Biol. Chem. 274:3055-3066(1999).
RN [6]
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RA TISSUE-Lymphoblast;
RX Hinds K., Sutterer C., Becker M., Hawkins M.;
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RA TISSUE-Lung;
RX Milgram S.L., Goldenring J.R., Schmidt P.H.;
RT "AKAP350: A multiply spliced family of proteins with centrosomal
RT association.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RA TISSUE-Brain;
RX MEDLINE-99087487; PubMed-9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [9]
RP SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.;
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [10]
RP FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND
CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
CC SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PPI)
CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.
CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.
CC SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND
CC CYTOSOL IN PARIENTAL CELLS.
CC ALTERNATIVE PRODUCTS: 6 ISOFORMS: 1 (SHOWN HERE), 2, 3/CG-
CC NAP, 4/YOTIAO, 5 AND 6/AKAP350, ARE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
CC DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC CAUTION: REF. 6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.
CC CAUTION: REF. 9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
DR EMBL: AJ131693; CAB40713.1; -;
DR EMBL: AB019691; BAA78718.1; -;
DR EMBL: AJ010770; CA009361.1; -;
DR EMBL: AF026245; AAB86384.1; -;
DR EMBL: AF083037; AAD22767.1; -;
DR EMBL: AC004013; AAB96867.1; -;
DR EMBL: AF091711; AAD39719.1; -;
DR EMBL: AB018346; BAA34523.1; -;
DR EMBL: AC000066; AAC60380.1; ALT_FRAME.
DR MIM: 604001; -;
KW Coiled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567
FT 164 914
FT DOMAIN 944 1022
FT 1100 1185
FT DOMAIN 1253 1280
FT 1336 1392
FT DOMAIN 1434 1459
FT 1585 1659
FT DOMAIN 1857 2435
FT 1857 2435
FT DOMAIN 2544 2561
FT 2603 2776
FT DOMAIN 3065 3092
FT 3124 3470
FT DOMAIN 3124 3470
FT 3587 3689
FT DOMAIN 3726 3730
FT 3726 3730
FT DOMAIN 203 292
FT 321 1011
FT DOMAIN 1846 2772
FT 17 28
FT VARSPLIC 1637 1642
FT VARSPLIC 1643 3911
FT VARSPLIC 2175 2183
FT VARSPLIC 2895 2907
FT VARSPLIC 2895 2948
FT VARSPLIC 3901 3911
FT VARIANT 1347 1347
FT 76 76
FT CONFLICT 475 475
FT CONFLICT 554 554
FT 638 638
FT CONFLICT 663 663
FT CONFLICT 913 913
FT CONFLICT 956 956
FT CONFLICT 980 982
FT 997 997
FT CONFLICT 1001 1001
FT CONFLICT 1020 1020
FT CONFLICT 1028 1028
FT CONFLICT 1626 1626
FT CONFLICT 1703 1703
FT 1707 1707
FT CONFLICT 1802 1803
FT CONFLICT 1843 1843
FT CONFLICT 1956 1956
FT CONFLICT 2027 2027
FT CONFLICT 2157 2158
FT 2169 2169
FT CONFLICT 2514 2514
FT CONFLICT 2851 2851
FT CONFLICT 2957 2957
FT CONFLICT 2983 2983
FT CONFLICT 3087 3087
FT CONFLICT 3218 3218
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -> D (IN REF. 5).
EI -> HE (IN REF. 7).
E -> V (IN REF. 3).
L -> R (IN REF. 3).
I -> N (IN REF. 8).
E -> D (IN REF. 3).
P -> S (IN REF. 3).
Q -> H (IN REF. 3).
Q -> H (IN REF. 3).
STTQFHAGMR -> ALSLTTSQHSAPTAPELITLSH
SLG (IN ISOFORM 6).
K -> KO.
/FTID=VAR_010926.
K -> KO.
R -> Q (IN REF. 3).
M -> I (IN REF. 3).
E -> G (IN REF. 3).
R -> S (IN REF. 3).
N -> S (IN REF. 3).
H -> N (IN REF. 3).
K -> N (IN REF. 3).
OKH -> PKP (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
Q -> P (IN REF. 1 AND 2).
N -> D (IN REF. 3).
V -> E (IN REF. 3).
R -> P (IN REF. 1 AND 2).
N -> T (IN REF. 3).
V -> G (IN REF. 3).
MISSING (IN REF. 5).
A -> P (IN REF. 3).
I -> V (IN REF. 3).
V -&

PRINTS: PRO0380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubules; ATP-binding; Coiled coil; Mitosis;
 Phosphorylation.
 FT DOMAIN 9 356 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).
 FT DOMAIN 357 479 COILED COIL (POTENTIAL).
 FT DOMAIN 479 729 COILED COIL (POTENTIAL).
 FT DOMAIN 729 895 COILED COIL (POTENTIAL).
 FT DOMAIN 895 1060 COILED COIL (POTENTIAL).
 FT NP_BIND 1033 1060 ATP (BY SIMILARITY).
 FT MOD_RES 930 930 PHOSPHORYLATION (BY CDC2) (BY
 SIMILARITY).
 SQ SEQUENCE 1060 AA; 119332 MW; 201E7EB6962B3492 CRC64;

Query Match 3.18; Score 164; DB 1; Length 1060;
 Best Local Similarity 20.38; Pred. No. 0.054;
 Matches 214; Conservative 155; Mismatches 383; Indels 302; Gaps 52;

24 TDLEFWNVNLRREVNICCEKEVDOAARGII---HMILKGSSECNLF---LKSLEW 76
 114 SDEFTW-----EODPLAGIIPRTLHQIEKLSKNGTEFSYKVSLEIY 137
 77 NYPLEFDLN---GQSL---FHQTSSEG---DLDDLAQDKD-LYH-----T 111
 158 NEELFPLSPSPVGRLOQFDDPRNKGVIKGLIEISVHNKDEYVHILDEGAARRKTA 217
 112 PSFLNFPYPLGED---IDILFKSTFTEPYLMKRDQHHHREQLTNGLLQALOSPCTIEG 169
 218 STLMNVYSSRSVSFVTLHMKETTYDVG-----EELVKGIGKMLVD---LAG 261
 170 ESKGKSTLLQRIAMLMGSGCKALTKFKVFPL---RLSRAQGLFETLDOLDIP 224
 262 SEWIGSGAVDKARERAGNINOSLILGRVITLALVERTPHIIPRESKLTPIIDSL---- 317
 225 GTIRKOTFAMM---LKTQRVLFLLDGYNEKP-ONCEPI-----EALIKENHFKN 273
 318 GGRKTSIIATVSPASINIEEV-STLDYANRAKSIMNKEVNOQLTKKLKE----- 370
 274 MVIYTTTECLRIHIOFGALTAEVG-----DMTEDSAQALLREVLKELAEDELLQIQK 327
 371 ---YTEELERLKRRLAAREKNVYLSENEQLOGKVLSEEMITETETKITAMEE 425
 328 SRCLRNIMKPTFVYITCA--IOMGESEFHSHTOTLTFHTFYLLIQKNKHKRGAASD 385
 426 LKISLSLFDONKKELEECTIIOCKEKEL-EETQNH-----QESKEQ---LAQES 472
 386 FIRSLDHCDDALEGVFSHKFDEFLQDVSSVNEVDLLTGLCKYTAQRPKPYKK-FHK 444
 473 FVVSARETEKRLHGT-ANKL-----LSTVRETRDVSGSLHKKLDRKKAVDHNFQVHE 525
 445 SPOEYTAGRLSLTSLTSHEPEYTKNGYLOKAV-SISDTTSYSSLARTCGSSVEATR 503
 526 NFAE-OMDRFESIQTIVDYISV-KOOGMLDPTNSIDDLGASSSRSLATAS----- 576
 504 AVAKHLAAYYOHCLLGLSLAKRPLRROESIOSVKNTEDEILKAININSFVCGIHLVY 563
 577 AVAKSPASV-----QETVTPQVSHSVSEIILKQETLS----- 607
 564 ESTSKSALSOEFAPFOGSLYINSNIP--DYLPDFEHLPMCASALDFI--KLDPYGG 619
 608 -SOAKGDLOOLMAAHRTGLEALRSDLLPVYTAVDLNLNLSHSLQNLFLVADKIDSHKE 666
 620 AMASWKAEDNGGIMMEAPETIYISRAVSLFFENKQOEETIE----- 663
 667 DMNSF--FTEHSHSLHKLRLDSSSLSSIQSEESLKEADITAOHMSHSEGVNVLISLON 724
 664 ---VTLRDS-----KLNK-----ODITYIG-----KIFSSATSLR 691
 725 QUNLGLMERTQQRSGFLSKGKRLQKSVGLQDDLDLVSSAETICISHHKKLAQESQDVA 764

QY 692 LQIKRAGVAGSLSLYSTCKNIYSLMEVNSPLTIDE-----RHITSVNLKLTLSIH 744
 DB VEIROLAG--SNMSTLEESKQCEKLTSSINTISQMCESAGKIDSVLEQCYLH 842
 QY 745 DLON--QRLPGLTDSIGN-----LKNLTKLMDINKNEEDAIKL 783
 DB 843 SSRKHQNLNLRKGVEDSGSSVEITRVNAQQAERKALTSLEQVRDDQEMVGEORLEL 902
 QY 784 ABEKLN-LKKMCFEHLTHL-SDIGEG-----MDYIVKSL--SSEPCDL-----EE 824
 DB 903 QEOVQSGLNKVVHSLKEELRLNDVPTGTPQRDPAAPSLVYTKRPDVLLEQFRQOOOEY 962
 QY 825 IOLVSCCLSANAVK-----ILAQNLHNLVSLTLDSENVLEKRG----- 864
 DB 963 LESISSVIS-EAVEPPEVQSLDEPPVAVNDVSIERSCDILSMTCQEKGGIRFPOKK 1021
 QY 865 -----NEALHELIDRMNVLEQ-----LRLALMP 887
 DB 1022 ALRKEKENRGNTTLERSKIMDEVQALTKSKLP 1055

RESULT 11

SPCA_HUMAN STANDARD; PRT; 2418 AA.
 ID P02349; Q15514;

AC 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE SPECTRIN ALPHA CHAIN, ERYTHROCYTE.
 GN SPTAI OR SPTA.

OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NCBI_TaxID=9606;
 RN [1]

RP MEDLINE=90170949; PubMed=1689726;
 RA Sahr K.E., Laurila P., Kotula L., Scarpa A.L., Coupal E., Leto T.L.,
 RA Linenbach A.J., Minkelmann J.C., Speicher D.W., Marchesi V.T.,
 RA Curtis P.J., Forget B.G.;

RT "The complete cDNA and polypeptide sequences of human erythroid
 alpha-spectrin.";

RT J. Biol. Chem. 265:4434-4443(1990).
 RN [2]

RP MEDLINE=9009318; PubMed=2794061;
 RA Sahr K.E., Tobe T., Scarpa A., Laughinghouse K., Marchesi S.L.,
 RA Agre P., Linenbach A.J., Marchesi V.T., Forget B.G.;

RT "Sequence and exon-intron organization of the DNA encoding the alpha
 I domain of human spectrin. Application to the study of mutations
 causing hereditary elliptocytosis.";

RT J. Clin. Invest. 84:1243-1252(1989).
 RN [3]

RP MEDLINE=84087888; PubMed=6654896;
 RA Speicher D.W., Davis G., Marchesi V.T.;

RT "Structure of human erythrocyte spectrin. II. The sequence of the
 alpha-I domain.";

RT J. Biol. Chem. 258:14938-14947(1983).
 RN [4]

RP MEDLINE=84087887; PubMed=6654895;
 RA Speicher D.W., Davis G., Yurchenco P.D., Marchesi V.T.;

RT "Structure of human erythrocyte spectrin. I. Isolation of the alpha-I
 domain and its cyanoen bromide peptides.";

RT J. Biol. Chem. 258:14931-14937(1983).
 RN [5]

RP MEDLINE=86205362; PubMed=3458204;
 RA Linenbach A.J., Speicher D.W., Marchesi V.T., Forget B.G.;

RT "Cloning of a portion of the chromosomal gene for human erythrocyte
 alpha-spectrin by using a synthetic gene fragment.";

RT Proc. Natl. Acad. Sci. U.S.A. 83:2397-2401(1986).
 RN [6]

RN [16]
 RP PARTIAL SEQUENCE.
 RX MEDLINE-84295638; PubMed-6472478;
 RA Speicher D.W., Marchesi V.T.;
 RT "Erythrocyte spectrin is comprised of many homologous triple helical
 segments.";
 RL Nature 311:177-180(1984).
 RN [17]
 RP IDENTIFICATION OF PROBABLE FRAMESHIFT IN 2407-2418.
 RA Gibson T.J.;
 RL Unpublished observations (MAR-1995).
 RN [18]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97001215; PubMed-8844207;
 RA Maillet P., Alloisio N., Morle L., Delaunay J.;
 RT "Spectrin mutations in hereditary elliptocytosis and hereditary
 spherocytosis.";
 RL Hum. Mutat. 8:97-107(1996).
 RN [19]
 RP VARIANT SER-24.
 RX MEDLINE-94289716; PubMed-8018926;
 RA Parquet N., Devaux I., Boulanger L., Galand C., Boivin P.,
 RL Lecomte M.-C., Dhermy D., Garbarg M.;
 RT "Identification of three novel spectrin alpha I/74 mutations in
 hereditary elliptocytosis: further support for a triple-stranded
 folding unit model of the spectrin heterodimer contact site.";
 RL Blood 84:303-308(1994).
 RN [10]
 RP VARIANTS CYS-28; HIS-28 AND SER-28.
 RX MEDLINE-91358728; PubMed-1679439;
 RA Coetzer T.L., Sahr K., Prechal J., Blacklock H., Peterson L., Koler R.,
 RL Doyle J., Manaster J., Palek J.;
 RT "Four different mutations in codon 28 of alpha spectrin are
 associated with structurally and functionally abnormal spectrin alpha
 I/74 in hereditary elliptocytosis.";
 RL J. Clin. Invest. 88:743-749(1991).
 RN [11]
 RP VARIANTS SER-28 AND ARG-48.
 RX MEDLINE-91346849; PubMed-1878597;
 RA Floyd P.B., Gallagher P.G., Valentini L.A., Davis M., Marchesi S.L.,
 RL Forget B.G.;
 RT "Heterogeneity of the molecular basis of hereditary
 pyropoikilocytosis and hereditary elliptocytosis associated with
 increased levels of the spectrin alpha I/74-kilodalton tryptic
 peptide.";
 RL Blood 78:1364-1372(1991).
 RN [12]
 RP VARIANT SER-45.
 RX MEDLINE-89323468; PubMed-2568862;
 RA Lecomte M.-C., Garbarg M., Grandchamp B., Feo C., Gautero H.,
 RL Devaux I., Bournier O., Galand C., D'Auriol L., Gilbert F.,
 RA Sahr K.E., Forget B.G., Boivin P., Dhermy D.;
 RT "Sp alpha I/78: a mutation of the alpha I spectrin domain in a white
 kindred with HE and HPP phenotypes.";
 RL Blood 74:1126-1133(1989).
 RN [13]
 RP VARIANT PRO-207.
 RX MEDLINE-92176375; PubMed-1541680;
 RA Gallagher P.G., Tse W.T., Coetzer T., Lecomte M.-C., Garbarg M.,
 RL Zarkowsky H.S., Baruchel A., Ballas S.K., Dhermy D., Palek J.,
 RA Forget B.G.;
 RT "A common type of the spectrin alpha I 46-50a-kD peptide abnormality
 in hereditary elliptocytosis and pyropoikilocytosis is associated
 with a mutation distant from the proteolytic cleavage site. Evidence
 for the functional importance of the triple helical model of
 spectrin.";
 RL J. Clin. Invest. 89:892-898(1992).
 RN [14]
 RP VARIANT VAL-1857.
 RX MEDLINE-9323053; PubMed-8486776;
 RA Wilmotte R., Marechal J., Morle L., Baklouti F., Philippe N.,
 RL Kastally R., Kotula L., Delaunay J., Alloisio N.;
 RT "Low expression allele alpha Ilevy of red cell spectrin is associated

RT with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and
 RT with partial skipping of exon 46.";
 RL J. Clin. Invest. 91:2091-2096(1993).
 RN [15]
 RP VARIANT BARCELONA.
 RX MEDLINE-93372367; PubMed-8364215;
 RA dalla Venezia N., Alloisio N., Forissier A., Denoroy L., Aymetich M.,
 RL Vives-Corrons J.L., Besalduch J., Besson I., Delaunay J.;
 RT "Elliptocytosis associated with the alpha 469 His->Pro
 mutation in spectrin Barcelona (alpha I/50-46b).";
 RL Blood 82:1661-1665(1993).
 RN [16]
 RP VARIANT CAGLIARI.
 RX MEDLINE-94043025; PubMed-8226774;
 RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,
 RL Jarolim P., Lorenzo F., Miraglia del Giudice E., Iolascon A.,
 RA Gallanelli R.;
 RT "Spectrin Cagliari: an Ala->Gly substitution in helix 1 of beta
 spectrin repeat 17 that severely disrupts the structure and self-
 association of the erythrocyte spectrin heterodimer.";
 RL J. Biol. Chem. 268:22656-22662(1993).
 RN [17]
 RP VARIANTS CULOZ AND LYON.
 RX MEDLINE-90347052; PubMed-2384601;
 RA Morle L., Roux A.-F., Alloisio N., Pothier B., Starck J., Denoroy J.,
 RL Morle F., Rudigoz R.-C., Forget B.G., Delaunay J., Godet J.;
 RT "Two elliptocytogenic alpha I/74 variants of the spectrin alpha I
 domain. Spectrin Culoz (G67->G77; alpha I 40 Gly->Val) and
 spectrin Lyon (C77->T77; alpha I 43 Leu->Phe).";
 RL J. Clin. Invest. 86:548-554(1990).
 RN [18]
 RP VARIANT JENDOURA.
 RX MEDLINE-92345619; PubMed-1638030;
 RA Alloisio N., Wilmotte R., Morle L., Baklouti F., Marechal J.,
 RL Dicluzeau M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,
 RA Delaunay J.;
 RT "Spectrin Jendouba: an alpha II/31 spectrin variant that is
 associated with elliptocytosis and carries a mutation distant from
 the dimer self-association site.";
 RL Blood 80:809-815(1992).
 RN [19]
 RP VARIANT TUNIS.
 RX MEDLINE-89323436; PubMed-2568861;
 RA Morle L., Morle F., Roux A.F., Godet J., Forget B.G., Denoroy L.,
 RL Garbarg M., Dhermy D., Kastally R., Delaunay J.;
 RT "Spectrin Tunis (Sp alpha I/78), an elliptocytogenic variant, is due
 to the C6G->T6G codon change (Arg->Trp) at position 35 of the
 alpha I domain.";
 RL Blood 74:828-832(1989).
 RN [20]
 RP VARIANT GENOVA.
 RX MEDLINE-94250920; PubMed-8193371;
 RA Perrotta S., del Giudice E.M., Alloisio N., Sclarratta G., Plauto L.,
 RL Delaunay J., Cuttillo S., Iolascon A.;
 RT "Mild elliptocytosis associated with the alpha 34 Arg->Trp mutation
 in spectrin Genova (alpha I/74).";
 RL Blood 83:3346-3349(1994).
 RN [21]
 RP VARIANT ANASTASIA.
 RX MEDLINE-95290423; PubMed-7772539;
 RA Perrotta S., Iolascon A., de Angelis F., Pagano L., Colonna G.,
 RL Cuttillo S., del Giudice E.M.;
 RT "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45
 Arg->Thr) with moderate elliptocytogenic potential.";
 RL Br. J. Haematol. 89:933-936(1995).
 CC -I- FUNCTION: SPECTRIN IS THE MAJOR CONSTITUENT OF THE CYTOSKELETAL
 CC NETWORK UNDERLYING THE ERYTHROCYTE PLASMA MEMBRANE. IT ASSOCIATES
 CC WITH BAND 4.1 AND ACTIN TO FORM THE CYTOSKELETAL SUPERSTRUCTURE OF
 CC THE ERYTHROCYTE PLASMA MEMBRANE.
 CC -I- SUBUNIT: COMPOSED OF NONHOMOLOGOUS CHAINS, ALPHA AND BETA, WHICH
 CC AGGREGATE SIDE-TO-SIDE IN AN ANTIPARALLEL FASHION TO FORM DIMERS,
 CC TETRAMERS, AND HIGHER POLYMERS.
 CC -I- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:

CC (1) N-TERMINAL DOMAIN (N),
 CC (2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
 CC (3) MIDDLE DOMAIN (M),
 CC (4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),

Query Match 2.9%; Score 152.5; DB 1; Length 2416;

Best Local Similarity 21.3%; Pred. No. 0.76; Matches 198; Conservative 128; Mismatches 319; Indels 285; Gaps 52;

176 STLLQRIAMLSGSKCAL---TKFKV-----FFLRISPAQGFPE---TLGDO-LLDI 223
 1472 ATRLQVLDNRKRLKQLIDERTKLDYANKQFYRDLELEWISMLPTACDESKDA 1531
 224 PGTIRF---QTFMAMLRKQRYFLLDGYNEFKPQNCPEIALIKE----- 267
 1532 TNIQRYLKQTFANHEVDGSEQVGVINLNGSLIERSCDNEAMKEQLEQKEMDHL 1591
 268 ---NIRFKNMVIVTTTECLRIHIFQALTA-----VGMTEDSA-----QAL 308
 1592 LERTNKGKRLNASHKQORENTSIRDFEWLSEAETLLAKDARLASAGNLLKHKHLL 1651
 309 IREVLIKE-----LAEGILL-----QIQSRCLRLMKTPFVVTTCALQMGESF 353
 1652 ERMMLAREDAKDLNTLAEDELLSSGTFNVDQIVKK--DNVKNRFL-----NVQELAA 1703
 354 FHSHTQ--TLFTFYDLLIQN--KHKHGVASDFINSLDHGDL-----ALEG-VFS 403
 1704 HHKRLKVAVALFOFQDLDDESMIEKLRVSSQDYGRDLQVQNLKHKRLKEGLVA 1763
 404 HKRDFLOVSSVNEVDVLLTGILCTYTAQRFKPKFKFHSQETAGRLSLTSHE 463
 1764 H--EPALQVNDL-----AEKLIDKAAVQOEEO-----LRLAQVEHWE 1801
 464 P--EEVAKGNG-----YLOKMWISD-----ITSTYSLRLTYCGSSVEATRA-VMKHL 509
 1802 KLKELAKAGIKLEESLEVIQFMQNAEBEBAWINEKNALAVRDCDDTLAATQSLMKHE 1861
 510 A-----AVT-----QHCCILGLSTAKRPLMRQESLOSQVKNTE-----QELKAININSF 554
 1862 ALENDRVAHETRYQVNCAGGEDILNKVLHESQNKIEISKIEALNKEPISLAKAIAMKL 1921
 555 VEGGILYQESISKALSQEFAPFOGKSLYI--NSGNIPDYLFDF----- 599
 1922 QLEDYDAFOGFMKADY--VEAMWADKETSLETNNGAD--LGDFTLLAKQDITLDASIQ 1977
 600 ---EHLPCASALDFIKL-----DEYGGAMASWEKAEDTGIMHEAPETY 643
 1978 SFGQERLPEITDLDK--KLISAGHNSKAIEERYAALLKRWEDLLE-ASAHHQKILEKQ 2034
 644 IP-SRAVSLFF-----NM--KOEFTLE---VILRDFSKLNKODITYLGI--- 683
 2035 LPLQKAEDELFEVFAHNASALNNWCEKMEENLSEPVCHVSLNEIROLOKHEDFLASARA 2094
 684 ---FSSATSRLQIKRCAGVAGS-----LSVLSTCKNYSIAME-ASPLTIEDERHIT 733
 2095 QADFKCLLEIDQOIK-ALGVPPSPYTWLYVEVETRWKHLSDIIEEREDLOKEEAROVK 2153
 734 SV-----TNKTLSDIHLQONR--LPQGLTDSLQNLKRLKILINDIKMNEDEAIKLA 784
 2154 NPEMCOEFQGNASTPLQWILETRAYFLDGLSLKETGTLE--SOLLEANKRKQKQIKQAMKR- 2210
 785 EGLKLNKKMCLFHLTLHSDIGEGMDYIVKSLSEPCDLEIQLVSCCLSANAVKIIAONL 844
 2211 ---OLTKIVLDGD-----NLEDAIILD----- 2229
 845 HNLVKLSILDSENYLEKNGNEALHELIDRM--NVLEQTLALMLPGKCDVQV-----S 895
 2230 ---IKSTIGLQDQW-----DQVQLGRLMOHNLQOIQAK-----DIKGVSEETLAE 2274
 896 LSLKLNHEEVPOLVRLGLKNNRLDTETIR 925
 2275 FSTYVYKHFE-----NLGRLTGHKEFR 2296

RESULT 12
 YMW6_YEAST
 ID YMW6_YEAST STANDARD: PRT: 1277 AA.

AC 004264: 004780:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 147.0 KDA PROTEIN IN ABF2-CHL12 INTERGENIC REGION.
 GN YMR076C OR YM9582.01C OR YM916.15C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 1-682 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (May-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 659-1227 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: 249259; CAA89222.1;
 CC DR EMBL: 248952; CAA88801.1;
 CC DR SGD: S0004681; YMR076C.
 KW Hypothetical protein.
 SQ SEQUENCE 1277 AA: 147040 MW: 9DF40A5274FD9623 CRC64;

Query Match 2.8%; Score 150; DB 1; Length 1277;

Best Local Similarity 18.8%; Pred. No. 0.44; Matches 183; Conservative 146; Mismatches 354; Indels 286; Gaps 41;

270 RKNWIVT---TTTECLRIHIFQALTAEGVGTEDSA-----QALIREVLIR 315
 10 KFNPSITISDQIISNELDLRLK--ALHHEELASIDQONTDLTGDXRYDALVSKLLK 66
 316 ELAEGILLQIQSKRLNLMK-----PLF--VITTCALQMGESFHSHTQTT 361
 67 HKDVG--IRAFTRACCLSDILRLVAPDAPYTDQNLDFIKLVLSQFEBOLDQENGHYHQOT 124
 362 ---LFTFYDLLIQNK-----HKHGVASDFIR 388
 125 YLTKLLEFRSYLLADLSSNNMLIELFHIFD-----FNKSPRALFNVIGIIGEVIS 180
 389 SLDHCGDLALLEGVSHKDFE---ELQDVSSVNEVDVLLTGL-LCKYTAQRF--PKY-- 439
 181 EFDVSPLLEVLRILFNKFLYNPNEIPEGIWNISDCGEVSLIDCPYVSNRMSRHLTKYVS 240
 440 KFFKHSFOETAGRLSSILTSHE-----PEVYTGNGYLQKMWV----- 479
 241 EITHETANDNNSRLTLVYVVKHLKVLRLMEVYVPELDINAVIGIYHLELSENELEFKKAT 300
 480 ---ISDTSYSSSLRYTCSS-----VEATRAVKKHLAAVYQHGCLLG 520
 301 KLIGQILLTISDINFEVSTHSDTFKAMISKIADISPDVRAWETESIPQITATEDISKELN 360
 521 LSIARK-----PLMRQESLOSQVKNTEQELKAININSFVEGIIHLYOESTK----- 568
 361 QALAKFTISDPRVTRTSVMEFKKVPVTEIKWNITKAIYTSILHLARKHKEVRELCIN 420
 569 ---SALSQEFAPFOGKSLYINSGNIPDYLF--FPEHLPCASALDFIKLDYGGAM 621


```

Db 684 NVLDSKSVETOK-LAYMELQOKAEFSQKHOKELIENMCKLT-----SOLAGOVE 733
QY 88 SLFH--QTSBGDDDLAODLKYHTPSFLNFPYPLGEDIDIFN-----LKSTFEPLV 139
Db 734 DLEHKQILSLNEMDKRCQODLHAEEESRLDLKSKDASLVNEDHQRBLLAFDQOPAM 793
QY 140 WRKQHHHRYEQLTLNGLLOALOSPCTIEGSGKSTLLO-RIAMWSSGCKALTKFK 198
Db 794 -----HHSFANIIGEOGSMPSERSECRLEADQSPKNSAILQNNVDLSL----- 835
QY 199 FVFFELSLRAGGLFELIDOLLIPGTIRKQFFMAMLLKRLQRYVFLDLYGNEFKPQNC 258
Db 836 -BESLSQKMSNDLQKQCELVQIGELIEN-----LMAEQM-----HOSFAETS 882
QY 259 PEEALIKENHFRKNVYVTTTECLNHLRQFALTAEVDMTEDSQAOLIREV-LIKEL 317
Db 883 QRSKIQEDISAQN-VVATLSALENKEKELELDLNDK-----ETFOAIOELKSNHL 936
QY 318 AEGLLIQKSRCLNLMKTPLEVVITCAIQMGSEFHSHTQ--TLEHTFYDL----- 369
Db 937 LEDSLLEQLLSEPLEKKEKMSIIS--LNRKEIEELTQENGTLKEINASINQEKMN 992
QY 370 LIOKNK-----HKHGVN-ASDFIRS-----LDHOGDL--ALEGVSHKPEFELQDVS 414
Db 993 LIOKSSFPANYIDEREKSISELSQYKQOEKLLIQCEETGNAYEDL-SQYK-AAQEK 1050
QY 415 SVNEDVLLTGLCKYTAQRFKRYKFFHKSFOEY-----TAGRRL 455
Db 1051 SKLECLINECTSLCENKKNKELEQLEKAFKHEQEFLLKLAFAERBNQNLMELEYQAL 1110
QY 456 SLLTSHHEPEVYKNGCYLOKMYISD---ITSTSSLLRYTCGSSVEVTRAVMKHLAA 511
Db 1111 RSEMTQNNNSKSEAGKQKEIMTLKEONKMOKEVNDLQEN-----EOLMKYK---- 1161
QY 512 VYQHGCLLGIKAPRLMROESLOS--VKNTTQOEILKAININSFVCGIHLQOES--TSK 568
Db 1162 -THHEC-----QNESEPIRSVAVERESEBNOCKFMKQMDLEVGEISLSDSY 1206
QY 569 SALSQFEAFQOKSLYXINGNIPDYLDFEHLNPNCAS-ALDEIKLDFYGAMASWEKA 627
Db 1207 NAQLOVLEALRNKELKLOES-----EKEKECLOHELOTIRDELTSNLQDMQ-- 1254
QY 628 AEDTGGIHME--APEYIIPS-RAVSLFFMWKQEF--TLEVTLRDPSKLKODITYIGKI 683
Db 1255 SQSISGLKQCEIDAEYEEYISGPHELSTSONDMAHLQCSLQTTMKNLELKEICELLQAEK 1314
QY 684 FSSATSLRLQIKRCAGVAGSLVLTCKNIVSLMVEASPL-----TIED----- 728
Db 1315 YELVTLENDRSRCITATRKMAEEVGKLNLEVKILINDSGLGELVEDLPGGEFGEPN 1374
QY 729 ERHITSVTNL-----KTLSTHDLQNRPLPGLTSLGINKLTKLMD----- 771
Db 1375 EQHPVSLAPIDESNSYEHLTSLDKEVQMHFALEQKFLSLQSEHKILHDHCQMSKMS 1434
QY 772 -----NIKMNEDDAIK-----LABGL-KNKKMCLFHLTHSDIGEGM 808
Db 1435 LQTYVDSLAKENLVLTNLTNLFQGLVAKEMQJGLEGLVPSLSSCCVPDSSSLSDSS 1494
QY 809 DY-----IYKSLSEPCDEEIOVSCCLSANAVK-----TLAOLNH-- 846
Db 1495 FYALALEQTDMSLNSLEBAVASANQSVDEVCSSLQTYVDSIKAKENLVLTNLRFPQ 1554
QY 847 -LVKLSILDSENYL-----EKDQNEALHELIDRMNVLEQDTALMLP 887
Db 1555 DLYKEMQJGLEGLVPSLSSCCVPDSSSLSDSSFYRALLBQTDGMSLNLGCVSA 1614
QY 888 WGDVQAGSLSSLLKH-----LEVPOLYKLGKMMWRLLDTETIRLGAFFGNPLKNFQOLN 943
Db 1615 NQCSVDEVEFCSSLOEBENLTREKTPSPAPAKVE-----FLESICEYR-----OSLE 1660
QY 944 LAGNRVSDGMLAFMGVFNJLQOLV-----PDSSTKEFLPDPALVAKLSOVLSKLFLQ 998
Db 1661 KLEEKHESQIKMKEL-QELEBOLLSERQELDLCKROYLSN-----EQMOOKLTSVTL 1716

```

```

QY 999 EARLVGWQFDDDDLSV 1014
Db 1717 ESKLAEKKQTEQLSL 1732

RESULT 14
BEM2_YEAST
ID BEM2_YEAST STANDARD; PRT; 2167 AA.
AC p39960;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GTPASE ACTIVATING PROTEIN BEM2/1P12.
OS BEM2 OR IP12 OR SUP9 OR YER155C.
OC Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=628BC;
RX MEDLINE=95050996; PubMed=7962097;
RA Kim Y., Francisco L., Chen G., Marcotte E., Chan C.S.;
RT "Control of cellular morphogenesis by the Ipi2/Bem2 GTPase-activating
RL protein: possible role of protein phosphorylation."
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050997; PubMed=7962098;
RA Peterson J., Zheng Y., Bender L., Myers A., Cerione R., Bender A.;
RT "Interactions between the bud emergence proteins Bem1p and Bem2p and
RL Rho-type GTPases in yeast."
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=628BC / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,
RA Hyman R., Kayser A., Komp C., Laskard D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=94123964; PubMed=8293973;
RA Chan C.S., Botstein D.;
RT "Isolation and characterization of chromosome-gain and increase-in-
RT ploidy mutants in yeast."
RL Genetics 135:677-691(1993).
CC -1- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR RHO1 AND RHO2.
CC INVOLVED IN THE CONTROL OF CELLULAR MORPHOGENESIS. REQUIRED FOR
CC PROPER BUD SITE SELECTION AND BUD EMERGENCE.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 GAP DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z5159; CA84524.1; -
CC DR EMBL: L33832; AAA57132.1; -
CC DR EMBL: U18917; AAB64682.1; -
CC DR SGD: S0000957; BEM2.
CC DR InterPro: IPR001849; PH.
CC DR InterPro: IPR000651; RASGEFN.

```

```

DR InterPro: IPR001895; RasGEF_CDC25.
DR InterPro: IPR000198; RhogAP.
DR Pfam: PF00169; PH. 1.
DR Pfam: PF00617; RasGEF. 1.
DR Pfam: PF00618; RasGEF. 1.
DR Pfam: PF00620; RhogAP. 1.
DR SMART: SM00233; PH. 1.
DR SMART: SM00234; RasGEF. 1.
DR SMART: SM00229; RasGEF. 1.
DR SMART: SM00324; RhogAP. 1.
DR PROSITE: PS50003; PH_DOMAIN. 1.
DR GTPase activation.
FT DOMAIN 16 25 POLY-SER.
FT DOMAIN 35 43 POLY-SER.
FT DOMAIN 58 63 POLY-HIS.
FT DOMAIN 198 208 POLY-ASN.
FT DOMAIN 253 260 POLY-SER.
FT DOMAIN 1161 1165 POLY-THR.
FT DOMAIN 1914 1948 PH.
FT DOMAIN 1981 2134 GAP DOMAIN.
SQ SEQUENCE 2167 AA; 245428 MW; DB7F4CD417E898F0 CRC64;

```

```

Query Match 2.8%; Score 148; DB 1; Length 2167;
Best Local Similarity 19.9%; Pred. No. 1.2;
Matches 205; Conservative 143; Mismatches 374; Indels 308; Gaps 53;

```

```

OY 6 DNSRALIQMGTVIKQITDLEF-VWN--VLNREVNITICCKVQDADARGITIMILKG 62
DB 1192 DSKKIMEDNTLVSSSLYDQNFVPMKMYVDENINILYMAKIQIGAAEALILHVKMY 1251
OY 63 S-----ESGN--LELKSLEK-----EW-----NYPLEFODLNGQ-----SLF 90
DB 1232 SDFTDLLCNSTLIDILIKMEDEVSTEMPTRIANSKLQSLPENVIETENLITLTDLF 1311
OY 91 H-----QTEGDDLDLAQ-----DLKDLVHTPSFLN----- 116
DB 1312 HGKSAVQKQLYRPIGVNRTQRTIDILNSFTSFDTLNNITDDPSDDMIRSFQKIH 1371
OY 117 -----FPLCEDIDIFNLKS-----TFPEVLMKRDQ-----HHHRVEQL 152
DB 1372 STNYEDILEWYIQIDNFIISKFNFLVSKRDWIVLFOLELLSKESLVSEFNPPLHFKSSKL 1431
OY 153 TLNGLLQALQ-----SPCIEGSGGKK-----STLLRIAMLGSGCKALTFKPF 199
DB 1432 INPGTLQHEFEISLFTWISLTLKDNGETSELEFEKLPQSTIKL-----IKLHTSLTT 1486
OY 200 VEFRLISRAOGLPE--TLCDQLDIPGTRKOTPMAMLLKQHVLELLDGYNEFKPON 257
DB 1487 FVFMISINWKNSSSERLTKCKVTLQILNVIK-----WKNSDLDFSEDESPHA 1536
OY 258 -CPLEALIK-----ENHRF-----KMWIVTTTECLR-----H 286
DB 1537 ICPHPIAFETIAHAHAIISPESRNYELSMIKASEKLSPTKCTQMLRSISNLEKIDIH 1596
OY 287 IROFALAEVGDMDTDSAOALI-----REVLIKELAEGL-----LQIQSRCLR 332
DB 1597 IKRF-----LEIDVFSKCKNLCPCPGWFIKSLLETISQVPMWSTTNSKLINFKRRFVN 1652
OY 333 NLMKTPLEFV-----ITCAIQMGSESEFHSHTQT-----LFHTFYDLLQKKN--HKHKG 380
DB 1653 NIIISVLDLIPEREPDIDEM--SDENPSKRTTGRIILNFNEDV-----KKVYKKTKK 1706
OY 381 VAASPFISLDHCGDLAEGVSHKFDLELDV--SSVNEVDYLLTGLLCKYTAQRFKPK 438
DB 1707 VSESAISE-----RFQROGVFNEILVNEIKIKREARKLELIDQEKILKNSSA----- 1756
OY 439 YKFFKSKOEYTAGRLSLTSHPEE-----VTRKNGVYLQKWSISDITTSYSLRLRT 494
DB 1757 ---LHQAVPK--KNKKSIVISGTHSDNDHSYINKNTGOTPSLGVME--SNNSRNRRD 1809
OY 495 CGSSVEATRA-----VWKHLAAYOHGCLLGLSLAKRPLMRQESLOSQVKNKTQOE 544

```

```

DB 1810 SRASEFTNRSSVSNSSHNCGKIKGCFRRPFSIGCFNTSSSNLSNLISQEVSSNKS 1869
OY 545 ILKAI--NINSVEGCHLHXOESTSKSALSOEFKFPQGSIXYNSGIPYLDLF----- 598
DB 1870 ILPSTLPEVDS-----KQIHDIKPSYSLKTEPKISIME-----IINHRNIPAVYFAKIVM 1920
OY 599 ---FEHLNFCASALD---FIKL-----DFYGMASMEKNAEDTGTGHHSEAPE---T 642
DB 1921 QNGHEYLQIOTASSDLTEWIKIRKASKRFSRSKRYK--KTHNKIFGVPLEDCEKENT 1978
OY 643 YIPSAVSLFENWKEFTEVTLDFESKLNKQDITTYLGKITFSATSRLQIKRCAGVAG 702
DB 1979 LIPTIVKLL-----EEIELRGIDEVG-----LYRIPSGIG 2009
OY 703 SISLVSLCKNINYSLM-VEASPLTIEDER--HITSVTLKTLSTHDLONQRLPGGLDLSL 759
DB 2010 SINAL---KNMFDEGATDMSFTLEDNRWEVNAIACCFKMYLIELPDSLFSHAMVDF 2065
OY 760 GNLLKULTKLIMDNIMKINEDAIKLAEGLNK-----KMCPLHL--THLSDIGECMD- 809
DB 2066 TDLAKIKYA---HAMVNEEYKRMNELLOKLPCTYQTLKRIVFLNKHVQHVANNKMDA 2122
OY 810 ---YIVKSL 816
DB 2123 SWLAIVFMS 2132

```

RESULT 15

ID	YBA4_YEAST	STANDARD	PRT	2493 AA
AC	P35194			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	HYPOTHEETICAL 287.5 KDA PROTEIN IN PDR3-HTA2 INTERGENIC REGION.			
GN	YBL0104W OR YBL0101.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RC	SEQUENCE OF 1-2066 FROM N.A.			
RA	STRAIN=S288C;			
RC	Rieger M.;			
RL	Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1214-2493 FROM N.A.			
RC	STRAIN=S288C;			
RX	MEDLINE=94378721; PubMed=8091860;			
RA	Wolfe K.H., Lohan A.J.E.;			
RT	"Sequence around the centromere of Saccharomyces cerevisiae			
RT	chromosome II: similarity of CEN2 to CEN4."			
RL	Yeast 10:541-546(1994).			
CC	-I- SIMILARITY: TO S.POMBE SPBC56F2.04.			
CC	-I- SIMILARITY: CONTAINS 4 HEAT REPEATS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z35765; CAAB84821.1; -			
DR	EMBL; Z26494; CAAB81266.1; -			
DR	PIR; S45734; S45734.			
DR	SGD; S00000100; YBL004W.			
KW	Hypothetical protein; Transmembrane; Repeat.			
FT	TRANSMEM 1170 1187 POTENTIAL.			
FT	DOMAIN 1480 1667 HEAT REPEATS DOMAIN.			
FT	CONFLICT 2440 2440 S -> R (IN REF. 2; CAAB84821).			
SQ	SEQUENCE 2493 AA; 287487 MW; A7EC4E3E9AE132B3 CRC64;			

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 10:55:45 ; Search time 35.13 Seconds
(without alignments)
4263.674 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 5281

Sequence: 1 MNEFKDSRALIQRMGMTVI.....WQFDDDLVITGAFKLVTVA 1024

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

rched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2244	42.5	439	4 Q9NPP4	Q9NPP4 homo sapien
2	296	5.6	1062	4 Q9Y239	Q9Y239 homo sapien
3	266	5.0	1062	4 Q9H6G6	Q9H6G6 homo sapien
4	264	5.0	1062	4 Q9NX02	Q9NX02 homo sapien
5	264	5.0	1062	4 Q9BVNS	Q9BVNS homo sapien
6	261.5	5.0	1033	4 Q9HAY9	Q9HAY9 homo sapien
7	242	4.6	1111	11 Q9JRM5	Q9JRM5 mus musculu
8	241	4.6	1111	11 Q9JLR2	Q9JLR2 mus musculu
9	234.5	4.4	1040	4 Q9HC29	Q9HC29 homo sapien
10	217	4.1	1429	4 Q9HAY8	Q9HAY8 homo sapien
11	212	4.0	1192	4 Q9UFT4	Q9UFT4 homo sapien
12	212	4.0	1429	4 Q9Y2E0	Q9Y2E0 homo sapien
13	212	4.0	1473	4 Q9C000	Q9C000 homo sapien
14	207.5	3.9	1399	4 Q9BZ29	Q9BZ29 homo sapien
15	207.5	3.9	1443	4 Q9BZ28	Q9BZ28 homo sapien
16	201	3.8	1031	10 Q9FHR9	Q9FHR9 arabidopsis
17	196.5	3.8	1590	10 Q9FTB5	Q9FTB5 arabidopsis
18	187	3.5	1405	10 Q50052	Q50052 arabidopsis
19	183	3.5	748	11 Q99MW0	Q99MW0 mus musculu

20	183	3.5	1187	10 Q9F114	Q9F114 arabidopsis
21	181	3.4	490	4 Q9H5N5	Q9H5N5 homo sapien
22	180.5	3.4	1054	7 Q78109	Q78109 mus musculu
23	180.5	3.4	1135	7 Q9TPE1	Q9TPE1 mus musculu
24	178.5	3.4	1031	10 Q9CAK0	Q9CAK0 arabidopsis
25	176.5	3.3	1108	10 Q9ARB2	Q9ARB2 linum usita
26	176.5	3.3	1108	10 Q9ARB0	Q9ARB0 linum usita
27	175	3.3	682	4 Q9NWM3	Q9NWM3 homo sapien
28	174.5	3.3	1398	10 Q9C784	Q9C784 arabidopsis
29	172.5	3.3	1560	10 Q9FTB6	Q9FTB6 arabidopsis
30	172	3.3	968	10 Q9ZTK1	Q9ZTK1 lycopersico
31	171.5	3.2	1110	10 Q9F177	Q9F177 arabidopsis
32	171.5	3.2	1167	10 Q9A471	Q9A471 arabidopsis
33	171.5	3.2	1295	10 Q9FNB3	Q9FNB3 arabidopsis
34	170.5	3.2	1220	10 Q24015	Q24015 lycopersico
35	170	3.2	1066	5 Q18902	Q18902 caenorhabdi
36	169.5	3.2	1110	10 Q9SDA5	Q9SDA5 arabidopsis
37	165	3.1	1112	10 Q41397	Q41397 lycopersico
38	165	3.1	1112	10 Q41398	Q41398 lycopersico
39	164.5	3.1	1715	10 Q49470	Q49470 arabidopsis
40	164	3.1	1016	10 Q9ZTJ6	Q9ZTJ6 lycopersico
41	163	3.1	512	11 Q9C076	Q9C076 mus musculu
42	163	3.1	657	11 Q9EP67	Q9EP67 mus musculu
43	163	3.1	853	10 Q9ZS82	Q9ZS82 lycopersico
44	162.5	3.1	524	4 Q9BTM6	Q9BTM6 homo sapien
45	162.5	3.1	1119	10 Q9ARB3	Q9ARB3 linum usita

ALIGNMENTS

RESULT	ID	Q9NPP4	PRELIMINARY:	PRT:	439 AA.
AC	Q9NPP4				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)				
DE	HYPOTHETICAL 49.3 KDA PROTEIN (FRAGMENT).				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Plummet R., Estivill X., Escarceller M., Sumoy L.;				
RA	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Auffray C., Ansorge W., Ballabio A., Estivill X., Gibson K.;				
RA	Lehrach H., Poustka A., Lundeberg U.;				
RT	"The European IMAGE Consortium for Integrated Molecular analysis of human gene transcripts.";				
RT	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.				
RL	EMBL: AL389934; CAB97523.1; -				
KW	Hypothetical protein.				
FT	NON_TER				
SO	SEQUENCE				
Query Match	42.5%; Score 2244; DB 4; Length 439;				
Best Local Similarity	99.8%; Pred. No. 4,4e-143;				
Matches	438; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	586	INSNIPDYLFDFEHLNFCASALDFIRLDFYGGAMASWERNAAEDTGGIHMEAEFTYIP	645		
DB	1	INSNIPDYLFDFEHLNFCASALDFIKLDFYGGAMASWERNAAEDTGGIHMEAEFTYIP	60		
QY	646	SRAVSLFFNNWKEQFTLEVTLRDFESKLKODITYIGKTFSSATSIRLDIKRCAGVAGLS	705		
DB	61	SRAVSLFFNNWKEQFTLEVTLRDFESKLKODIRYIGKTFSSATSIRLDIKRCAGVAGLS	120		
QY	706	LVLSTCKNIYSIMVASEPLTIEDERHITSVNLKTLSTHLDONQLPGGLTDSLGKLNKL	765		

```
|||||
Db 121 LVYSTCKNIYSIMVEASPLTIEDERHITSVTNKLTSIHDLQNRPLGGLTDLGNLKNL 180
OY 766 TKLIMONIKMNEEDAIKILAGLNKMKCFHLTHLSDIGEGMDYIVKSLSSPCDLEI 825
Db 181 TKLIMONIKMNEEDAIKILAGLNKMKCFHLTHLSDIGEGMDYIVKSLSSPCDLEI 240
OY 826 QLVSCCLSANAVKILAQNLHNLVLSILDSENYLEKDGNEALHELIDRMNVLEQTLALM 885
Db 241 QLVSCCLSANAVKILAQNLHNLVLSILDSENYLEKDGNEALHELIDRMNVLEQTLALM 300
OY 886 LPMGCVQVGLSSLLKHLLEVPOLYKIGLNKMRLTDEIHLGAFGKNPLKNFQDLNLA 945
Db 301 LPMGCVQVGLSSLLKHLLEVPOLYKIGLNKMRLTDEIHLGAFGKNPLKNFQDLNLA 360
OY 946 GNRVSSDGMVLAFGVFNELKQVFEFDEFTKEFLPDPALVKLSOVLSKLTFLQEARLVGM 1005
Db 361 GNRVSSDGMVLAFGVFNELKQVFEFDEFTKEFLPDPALVKLSOVLSKLTFLQEARLVGM 420
OY 1006 QFDDDLVITGAFKLVTVA 1024
421 QFDDDLVITGAFKLVTVA 439
```

```
RESULT 2
OY1239 PRELIMINARY: PRT: 953 AA.
AC OY2239;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NOD1 PROTEIN.
GN NOD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99262599; PubMed-10329646;
RA Inohara N., Koseki T., del Peso L., Hu Y., Yee C., Chen S., Carrio R.,
RA Merino J., Liu D., Ni J., Nunez G.;
RT "Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-
RT kappaB."
RL J. Biol. Chem. 274:14560-14567(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99240667; PubMed-10224040;
RA Bertin J., Nir W.-J., Fischer C.M., Tayber O.V., Errada P.R.,
RA Grant J.R., Kelly J.J., Gosselin M.L., Robison K.E., Wong G.H.W.,
RA Glucksmann M.A., Distefero P.S.;
RA "Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member
RA that activates NF-kappaB."
RL J. Biol. Chem. 274:12955-12958(1999).
DR EMBL: AF149774; AAD43922.1; -
DR EMBL: AF11925; AAD28350.1; -
DR EMBL: AF126484; AAD29125.1; -
DR InterPro: IPR001315; CARD.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00619; CARD; 1.
DR PROSITE: PS50209; CARD; 1.
DR SMART: SM00114; CARD; 1.
DR SMART: SM00370; LRR; 5.
DR SEQUENCE 953 AA; 107690 MM; 0A9DF5FC6487E21A CRC64;
```

Query Match 5.6%; Score 296; DB 4; Length 953;

Best Local Similarity 20.8%; Pred. No. 2e-11; Mismatches 356; Indels 406; Gaps 58;

```
OY 1 MNFKNSRLAIQRMGTIVKQITDLEFVNVNVRREVNICEKVEQDARGIHHILK 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 IOLKSNRELVTIHRNT-QCLVDNLKNDYFSAEDAEIVCACPTPOPKVKILDILOS 77
```

```
OY 61 KGSSECNLFKSLKENNYPFLFODLQSLFHOQSEGLDLDLADLK----DIYHPSPFLN 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 KGESEVSEFLYLQ-----LADAVYDLRPMLEIGFSPSLT 115
OY 117 FYPLGEDIDIIIFNLKSTFEPV-LMRKDOHH-----RYEQLTL----- 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 -----QSKVVVN-----DDPVSRYYQOLRHHGLRDSKEVLCAOKEEILLEIYMDITM 164
OY 155 -----NGLLQALQS-PCITIE-----GSGKGSKSTLQRIAMWGSGKC 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 ELVGFSENESTGSLNSLACLDDHTTGTLNEOGETIFLIGAGVKSMLDRDLSMATGYL 224
OY 192 KALTFRKFFVEFLRSRA-----OGGLFETLCDDLDIPGTIRKQTFMMLTKLROR 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 DAVGVKEFFHRCGMFCFKESDRCLQDILLFKHYC-----YPERDEEVF-AFLIRFHY 278
OY 243 VFLFDGYNF-----KPNCEPEIALIKENHRKNMYIVTT--TTECLRHRQFG 291
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 ALFTFDGLDELHSDLDLSRVPDSSCPWEPA-----HPLVILANLISGLKLK-----G 325
OY 292 A--LTAEGVDMTDESAQALIREVL-----KELAGLLIQOKSRC 330
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 ASKLITARTG--IEVPRQPLRKVLLRGFSPSHLRAYARMPEERLQRLLSQLEARN 383
OY 331 LRNLKPTPLF--VVITCAIQMGSEFEHSHQ-----TTFHFHY-----DLIT 371
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 LCLSLCVPLFCWIIIFRC-FQHFRFAFEGSPQDPDCTMTFLTVDVLTVEHLNMQSSIV 442
OY 372 QKRNKR-----IKGVAASDFIRSLDLCGLALEGVSHKFDPELOV--SSVNE-DVLLT 423
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 QNRTSRSPVETLHAG--RDTLCSL--GOVAHRGMKSLFVFTQEEVQASGLQERMOY- 495
OY 424 TGLCKVYTAQRFKPK-----YKFFH-----KSPQEX- 449
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 -GPL-----RALPELCPGDDQSYEFPHLTQAFATFVLDVADVQOELLRFQEMM 548
OY 450 -TAGRLSLITSHPEEYVTKGNGYLQKNVYSIDITSTYSSLLRYCGSSVEATRAVMKH 508
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 PPAGAAATTSCYPPFLPFQCGSGSPAREDFLNKHDFQFTNL--FLCGLLSKAKQKILRH 606
OY 509 LAAVYQHGLLGSLIAKRPIMRW--ESLOSVKNTTQOELLKAININSF----- 554
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 LVPA-----AALRRKRKALMAHLESSLKRGYLS-----LPRQVSEFNOVQOAMPTTMM 655
OY 555 VECGHLVQESTSKALSQEFQKSLYINSGNIPLYLDFEFHPLNC-----ASA 608
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 LRC---IYEQSQKVG--QLAA-----RGICANVL-----KLTYCNAACSADCSA 694
OY 609 LDFIKLDFYGCAMASWEKAEDTGTGIMHEBAPETIIPSR-AVSLFRNMKQOEFRTLEVTLR 667
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 695 LSFV-----LH-----HPPKRLADL----- 710
OY 668 DFKLKKQDITVYLGKTFESSATSLRIQKCA-GVAASLSYVSTCKNIYSIMVEASPLTI 726
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 DNNNLNDYGVRELQPCPSRLTYLRLSVNQITDGGKVLSELTQKIVYIYVGLYNQITD 770
OY 727 EDERHITSYTNLKTLSIHDLQNRPLGGLTDLGNLKNLTKLIMDKIKMNEEDAIKILAG 786
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 VGARVYTKI-----LDECKGLTHLKGKNTLSBSGKTLALA 807
OY 787 LKNLKKMKCFHLTHLSDIG-----EGMDYIVKSLSSPCDLEIQLVSCCLSANAYK 838
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 VKNSKS-----ISEVGMGNQVGDGAKAFALRNHP-SLTJTLSTASNGISTEGCG 858
OY 839 ILAQNLHNLVKSILDSSENYLEKDGNEALHELIDRMNVLEQTLALMPGCVQVQSSLS 898
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 859 STARALQONTSLIIMLTQNLDEVAESLAEML-----KVNO 896
OY 899 LKHLLEVPOLVYKIGLNKMRLTDEIRIILGAFGKNPLKN---FOQDLNAGNRVSSDGM 955
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 897 TKHL-----W-LIQNTITAKTQALDALALQSNIGTIECLNGNLIRFEE-- 940
```



```

QY 956 AFMGVFNKQLVFP 970
DB 941 --AKVYEDEKRITCF 953

RESULT 3
Q9H6G6 PRELIMINARY; PRT; 1062 AA.
AC Q9H6G6:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CDNA: FLJ22299 FIS, CLONE HRC04653.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;

[1]
SEQUENCE FROM N.A.
MA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isigai T., Sugano S.,
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025952; BAB15293.1;
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR003590; LRR_out.
DR InterPro; IPR003592; Disease_resist.
DR PRINTS; PR00364; DISEASERIST.
DR SMART; SM00370; LRR; 7.
DR SMART; SM00368; LRR_R1; 8.
SQ SEQUENCE 1062 AA; 120488 MW; 9F890DF9B7CB944 CRC64;

Query Match 5.0%; Score 266; DB 4; Length 1062;
Best Local Similarity 22.5%; Pred. No. 2.5e-09;
Matches 201; Conservative 123; Mismatches 337; Indels 234; Gaps 42;

QY 166 IIEGSGKSTLLQRIAMLMGSKKALTRKFEVFL--RLSRAQGLFETLDDLLD 222
DB 210 VLYGAGIGKTTLLQKIMLDAED--NLHKKYAFYLSCRELSRLGPCSPAEIV--FRD 265
QY 223 IPRTRKOTFMALKLQRYVFLDGYNEFK-----PONCPREI-----ALI 265
DB 266 WP---ELDDIPHLIAQARKILFVIDGFDELGAARGALIEDSCGMKKKPPVVLGSL 322
QY 266 KENHFKMNVYITTECLRIHFGALTAE-----VGMTEDSAQALIREVLIKELA 318
DB 333 NKVMLPKAALLVTRPRALRLDR---LAEERPIYIRVEGFLEEDRRAYFLRHFGDEDOA 378
QY 319 EGLLQIQKSRCLRLMLTPLFVVITCA---IOMGESEFHSHTOTLTHTFYDLLQKNK 375
DB 379 MRAFLMSNAALFOLGSAPVACWIVCTTLKLOMEKGEDEVPTCLTRGLFLRFKSRP 438
QY 376 HKHKAASDFRSLDHGDLALGCVFS---HKFDELQDVSSVNEVDVLTITGLCKY 430
DB 439 ---QGAQRLGALRTLSL---LAAQGLMAQTVSLHREDLERLGVOGSDRLRFIDGILQOD 492
QY 431 TAQRKPKYKFFHKSFQETYAGRLSLTSHREPEVTGKNGYLOKMWISIDITSTYSSL 490
DB 493 RVS--KGCYSFTHLSFOOF---LTAFLYLTLEKEEEDRQHTW---DIGVQKLISGV 542
QY 491 LR-----YTCGSSVETRAVMKHLAAVYOHGCLLGISTAKRPLMR----- 530
DB 543 ERLRNPDLQAGYISFGLANE-KRA--KELEATF--GGRMSPTI-KQELRLDISCKGSH 596
QY 531 -----QESLQSVKNTTDEILKAININSFVCGIHLVQESTSKSALQEFAPFOGKSL 584
DB 597 STVTDLOELGLGYESOEELVKYV-MAQFKETSLHNAVDVPPSPFCVKKHCRNLQKNSL 655
QY 585 YINSNINID-----YLPDFEHLPCA---SALDPIKL---DPFGG 619
DB 656 QVIRENLEPENTVASEDAEVERSODDOHMLPWTDL--CSIFGSKNDLMGLAINDSFLSA 713

```

```

QY 620 AMAS--WEKAEDTGSIHWEAPETIYIPSRAYSLEFNKMQERTLEVTLRDSKTL----- 672
DB 714 SLVRLCEQIASDT--CHLQ-----RVYFKNISPADAHRLCLLRGHKTYTLTL 762
QY 673 --NKOD-----ITYLGKIFSSATSLRLQIKRAGVAGSLVLS----- 709
DB 763 QGNDDQDMFPALCEVLRHPECLRYLGLVSCSATYQW-----ADLSIALEVNOSL 813
QY 710 TCKNT--YSLWVASPLTIEDERHITSYTNLKTLSIHDLQNRPLRGLTDSIGNLKNLTK 767
DB 814 TCVNLSDELDEGAKLLYTLTRHKCF-----LQR 844
QY 768 LIMDNINKNEEDPAIKLAEGLKMLKMCFLHTL-----SDIGE--GMQYIVKSLSEPPDL 822
DB 845 LSLENCHLLEANKCDLAAVLVVSRE---LTHLCLAKNPIGNTGKFLCBGLRPRECKL 899
QY 823 BEIQLVSCLSANAVKIIAONLHNLVKLSIDLSENVLEKDGNEALHBLDRMNYLEQTL 882
DB 900 QTLVLMNCDITSDGCCDLTKLLQEKSSLLCDLGLNHIGVGMKFLCALRK--PLCNLRL 957
QY 883 ALMLPWCDDVQG--SLSSLLKHLLEVPLVKGLKMMRLTDEIRILGAFPGKNPL 936
DB 958 CLML-WGCSIPPFSCEDLCSALSCNSQSLVTIDL-----CONPL 994

RESULT 4
Q9NX02 PRELIMINARY; PRT; 1062 AA.
AC Q9NX02:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NALP2 (HYPOTHETICAL, 120.5 KDA PROTEIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;

[1]
SEQUENCE FROM N.A.
MA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isigai T., Sugano S.,
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Martignon F., Hofmann K., Tschopp J.;
RT "NALP2 a novel NACHT, LRR and PYD containing protein."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000517; BAA91223.1;
DR EMBL; AF310106; AAG30289.1;
DR EMBL; BC003592; AAH03592.1;
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR003592; LRR_out.
DR PRINTS; PR00364; DISEASERIST.
DR SMART; SM00370; LRR; 7.
KW Hypothetical protein.
SQ SEQUENCE 1062 AA; 120514 MW; 4DBB0F6E9C2BC8A7 CRC64;

Query Match 5.0%; Score 264; DB 4; Length 1062;
Best Local Similarity 22.5%; Pred. No. 3.4e-09;
Matches 201; Conservative 123; Mismatches 337; Indels 234; Gaps 42;

QY 166 IIEGSGKSTLLQRIAMLMGSKKALTRKFEVFL--RLSRAQGLFETLDDLLD 222
DB 210 VLYGAGIGKTTLLQKIMLDAED--NLHKKYAFYLSCRELSRLGPCSPAEIV--FRD 265

```

QY	223	IPGTRKQTFAMALTKLRQVRLFLDDQNEF-----KPNCEP-IEALI	265
Db	266	WP-----ELIADDIPLHIAQARKLTLEFVIDGFDEGAAPGALIEDICGDEWKKKRVVYLSSLT	322
QY	266	KENHRKFMVIVTTTTRBCLRHIRIOFGALTAE-----VGDMDSDAQAIIREVLIKELA	318
Db	323	NRVMLPKAALLVTRPRALRDLR-----ILAEPIYIRVEGFLIEDBRAYFLRHGDEDDQA	378
QY	319	EGLLLOIQKSCGLNLMKTPLEFVYITCA-----IQWGESEFHSHTQTLFHPHYDLLIOKN	375
Db	379	MRAFELKMSNALRQLDSAPAVCMIVCTTTLKLOMEKEKEDVPVLTTLTGFLRFLCSRF	438
QY	376	HKHKVAAADPFIKSLDHCGLDALGVS-----HKPFELQDVSVAVEDVLTLLGLCKY	430
Db	439	---OGAQLRGALRLSL-----LAAQGLWAQISVYLRHREDLERGAVESDLRLFLDGDILRQD	422
QY	431	TAQRFKPKYKFEFHKSFOEYTAGRRLLSLTSHSEDEVTKGNGYLQKKVMSISDIITYVSSL	490
Db	493	RVS---KCGYSFIHLSFOQF-----LFALEFYLKEKEEEDRGHTW---DIGDQKLLSGV	542
	491	LR-----YTCGSSVATPAVAKKHIAAYQHCCLGLSLAKRPYMR-----	530
QY	543	ERLRNPDLIQAGYVSFGLANE-KKA--KELEATP--GCRMSPD-LKQELLRCDISCKGH	596
Db	531	-----GESLQSVKNTTEOEILKAININSFEGCIHLYOESTSKSAISOEFEEAFQGS	584
Db	597	STVTDLQELLOCLIVESQEEELYKVEV-MAQFEIISHLNANDVAVSSCVNHCNRLQKMSL	655
QY	585	YINGSNIPD-----YLFDFEFHLPNCA---SALDFIKL---DFYGG	619
Db	656	QVIRENLPENVTASESDAEVRSODDOHMLPFMTDL--CSIFGSKDLMGLAINDSFLSA	713
QY	620	AMAS--WEKAAEDGGIHMEAPRTYIIPSRVSLFFMMKQEFRLTEVTLRDFSKL-----	672
Db	714	SLVRLCEQIASDT--CHLO-----RVYFNINISPADARRNLCIALNGHKTIVYTL	762
QY	673	---NKOD-----ITYLGKIFSSATSLRLOIKRCGAVAGSISVLYS-----	709
Db	763	QGNDDQDMFALCEVLRHPECNLRGLGVSCSANTQOM-----ADSLALEVQSL	813
QY	710	TCKNI--YSLMVEASPLTIEDERHITSYTNLKTLSIHDLQNRLLPGLTJSLGKALKLTK	767
Db	814	TCVNLSDNELLEDGAKLYTLRLHRCGF-----LQR	844
QY	768	LIMONIKNNEEDAIKLABGLKNLKKMCLFHLTHL-----SDIGE-GMDYIVKSSSEPCDL	822
Db	845	LSLEKCHLTFEANCDDLAAVLVYSRE-----LTHCLAKNPIGNTGAVFLDEGLRYPECKL	899
QY	823	EEIOLVSCSLSANAVKILIAONLHNLVKLSTLSDSENYLEKDGNEALHELIDRNVNEQOLT	882
Db	900	QTLVLMNCIDITSDDCCDCTLKLOEKSSSLDLDLGNHIGVGMKFLCEALRK--PLCNLR	957
QY	883	ALMPWCGDVQG-SLSSLLKHLLEVPQYKLGKLNKWRKLTDEIRIILCAFGKKNL	936
Db	958	CLML-WGCSIPFSCEDLCSALSCNQSILVTLDL-----GQNP	994
RESULT	5		
Q9BYNS			
AC	Q9BYNS	PRELIMINARY;	PRT; 1062 AA.
DT	01-JUN-2001 (TREMblrel, 17, Created)		
DT	01-JUN-2001 (TREMblrel, 17, Last sequence update)		
DT	01-JUN-2001 (TREMblrel, 17, Last annotation update)		
DE	HYPOHETICAL 120.6 KDA PROTEIN.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
CC	TISSUE=CHORIOCARCINOMA.		

RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases
DR EMBL: BC001039; AAH01039.1; -
KW Hypothetical protein.
SQ SEQUENCE 1062 AA; 120572 MW; 4DBB0A7A9C2BC8A7 CRC64;

Query Match	5.0%;	Score 264;	DB 4;	Length 1062;
Best Local Similarity	22.5%;	Pred. No. 3.4e-09;		
Matches 201;	Conservative 123;	Mismatches 337;	Indels 234;	Gaps 42

QY	106	11BESGNSSTLQKRAMMGSSKALNKRVFL---	KLSKAGGJFLFDQJLD	227
Dd	210	VLYPAGLGTTLQAKIMLMDAED--NLHKFYAFYLS	CRELSRLQPCSFAEIYV--FRD	265
QY	223	IPGIRKOTFAMMLKLQRVTLFDLDYNEF-----	KRONCE-IEALI	265
Dd	266	WP---ELQDDIPHLAQAARKLFLFYIDGFD	EAGPGLIEDICGDMWKRPVPLVGLSL	322
QY	266	KENHFKNMVIVTTECTLRHIFRQALTAE-----	VEDMTEDSAQALIREVLKELA	318
Dd	323	NRVMLPRKALLVTLTPRALRNLR---ILABEPIY	INVEGFLEDRRAYLRLHFGDEDA	378
QY	319	EGLLLOJQKSRLNMLKTPLEVVITCA---IOMGE	SEFHSQOTTLFHTFYDLLQKNK	375
Dd	379	MRAFELKRSNAALFOLGSAPAVCIYVETTLK	QMEKEDVPRLCTPRTGFLFRGLSRFP	438
QY	376	HKHKGAASDFIRSLDHCGLDALEGVSS-----	HKPFELQDVSVMEDVLLTGLLCKY	430
Dd	439	---OGAQLRGLRLSL---LAAGLMAQISVLR	REDLERLQVQESDLRLFLGGDILROD	432
QY	431	TAORFKRYKFKFHKSFOEYTAGRRLSILTS	SHEPEVTNGVYLQKKVSIISDITSVYSL	490
Dd	493	RVS--KGYSPFIHLSFOGF-----L7RAL	FTLEKEEBEDBDGHTW--DIGDVKILLGV	542
QY	491	LR-----YTGSSVEATRAVYKHNLAAYQH	CCGLSLSTAKRPLR---	530
Dd	543	ERLRNPDLLQAGVYSFGLANE-KRA--KELEATP-	-GCRSPRI-KOELLRCDISCKGH	596
QY	531	-----OESLOSANTYTHEOELIKININSPE	GITHYORSTKSALSOEFMAFQOGLK	584
Dd	597	STVLDLOELLGLCLYESQEBELVKEV--MAQK	EISLHNAVDVPPSCVYKHCNRLQKMSL	655
QY	585	YINSGNIRP-----YLFDEFENLRPNCA---	SALDFIKL---DFYGG	619
Dd	656	QVIKENIPENTVASESDAEBERSQDDQHM	IRPFWIDL--CSIFGSNKDMLGALINDSPLSA	713
QY	620	AMAS--WEKAEDTGGIHMERAPETYIPSAVSI	LFENWKQEFRTLEVTLRDESKL-----	672
Dd	714	SLVRILCEQIANSPT--CHLO-----RVR	KNISPADHNRMLCALRHKTYVTLTL	762
QY	673	---NRQD-----ITYLGIFSSATSLSLQIK	RQAGVAGSVSILVUS-----	709
Dd	763	QGNQDDMFRLCLCVLRLHPRCNLRYGLCV	SCATQOM-----ADSLALEVQSL	813
QY	710	TCXNI--YSIMVEASPLTIEDERHITSVNLK	TLSTHDLQNRLPGLTDSLGNLNLK	767
Dd	814	TCVMLSQNNELLDECAKLTYTLTLRHPKF-----	---	844
QY	768	LIMNITMNEDEAIKLAEBGLKNLKMOLFHLTHL---	SDIGE-GMDYIKSLSSEPCDL	822
Dd	845	LSTENCHLTEANCIDLAAVLPVSRK-----L	THCLAKNPIGNTGVKFLCEGLRYEBCKL	899
QY	823	EEIOLVSCCISANAVKILIAONLHLVYLSL	DSYNTSENYLEKDGDEALHELIDRNNVLEOLT	882
Dd	900	QTVLVMKCDITSDCCDLTKILDEKSSLLC	DLGLNHIYVKGAKFLCEALRK--PLCNLR	957
QY	883	ALMLPMGCDVQG--SLSSILKHLBEVPLVYK	LGIKNMRLDTETIRILGAFEGKKNPL	936
Dd	958	CLML-WGCSIPFPFCEDLCSALSCNOSL	VTLDL-----GQNP	994

Q9HAV9 PRELIMINARY; PRT: 1033 AA.
 ID Q9HAV9
 AC Q9HAV9
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NUCLEOTIDE-BINDING SITE PROTEIN 1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bertin J., Distefano P.S.;
 RT "The PRIN Domain: A Novel Motif found in Apoptosis and Inflammation
 proteins.";
 Cell Death Differ. 0:0-0(2000).
 EMBL; AF298547; AAG15253.1;
 InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003590; LRR_Rninh.
 DR PRINTS; PR00364; DISEASERLIST.
 DR SMART; SM00370; LRR_6.
 DR SMART; SM00368; LRR_R1; 8.
 SQ SEQUENCE 1033 AA; 117387 MW; C16152AD510BCA63 CRC64;

Query Match 5.0%; Score 261.5; DB 4; Length 1033;
 Best Local Similarity 22.5%; Pred. No. 4.8e-09;
 Matches 201; Conservative 123; Mismatches 336; Indels 235; Gaps 43;

166 IIEGSGKSTLLQRIAMWGSGKALTKPKFVFLL---RLSRAOGLEFETLCDQLLD 222
 DB 182 VLYGKAGIGKTTLAQKMLMDNAED--NIHKFYAFYLSCHRELSLGGCSFAELY--FRD 237
 QY 223 IPGTRKOTFAMMLKLRLQRYVLELDGYNEF-----KPNCPK-IEALI 265
 DB 238 WP---ELDDIPHILAQARKILFVDFGDELGAARGALIEDICGDMKKRPVVLGSL 294
 QY 266 KENHFKMNVYTTTTECLRNHRGALTAE-----VGMTESSAALLREVILKELA 318
 DB 295 NKVMLPKAALLVTTTPRALRLDR---ILAEPIYIRVEGLEEDKRAYFLNHGDEDOA 350
 QY 319 ECLLIQIOKSRCLRMKTPLEVVITCA---IOMGESEFHSHTQTLTHTEFYDLTIQNK 375
 DB 351 MRAFLMMSNAALFOLGSAAPAVCAIVCTTKLQMEKGEDVPVPTCTRGCLFRLCSKRP 410
 QY 376 HKHKGVAADFIKSLDHCGLALEGVFS---HKFDELODVSSVNDVLLTGLLCKY 430
 DB 411 ---QGAQIRGALRTLSTL---LAAQGLMAQTVSLHREDLERLGVQSDRLRLFDGDI 464
 QY 431 TAQRKPKYKFFHKSFQETYACRLSLTSHPEEVTKGNGYLOKMSISDITSTYSSL 490
 DB 465 RVS--KGVSYFHLSPFOF---LTALEPYLKEEEDRDGHTW---DIGVQKLSLV 514
 QY 491 LR-----YCGSSVEATRAVMKHLAAVYOHGCLLGLSIAKRPMLR----- 530
 DB 515 EKLRLPDLIOAGYISFGLANE-KRA--KELEATF--GCRMSDIT-KQELRCDISCKG 568
 QY 531 -----QESLOSVKNTTQELIKAININSFVCGIHLVQESTSKALSQEFAPFOGSK 584
 DB 569 STVLOLQELGLCYSEBELVKEV--MAQFKRISLHLNAVDPVPSFCVKKHCRNLQKMS 627
 QY 585 YINSNRPD-----YLDFEHLPMCA---SALDFIKL---DFYGG 619
 DB 620 AMAA--WEKAADTGCIHMEAPETIYIPRAVSLFNNMKOEPTLEVTLRDSKSL----- 672
 DB 666 SLVRLICEIDIASDT--CHIQ-----RYVFKNISPADAHNCLALRGKITYTYLT 734
 QY 673 --NKOD-----ITVLGKIFSSATSLRLQIKRCAGVAGSLSLVLS----- 709

DB 735 QGNDQDMFPALECEVLRHPECNLRYLGLVSCSAATQGW-----ADLSIALEVNSL 785
 QY 710 TCKNI--YSLWEASPLTIEDERHITSVTNKTISTIHDLQNRPLGGLTSLGNKMLTK 767
 DB 786 TCVNISDNELLDEGAKKLYTLTRHKCP-----LQR 816
 QY 768 LIMDNKKNNEEDAIKLEGLKNKKMCLFHLTHL-----SDICE-GMDYIVKLSSEPCDL 822
 DB 817 LSLKCHLTENCMCDLAAVLVRSR-----LTHICLAKNPFGNNGVFKLGLGAYPECKL 871
 QY 823 EBIOLVSCCLSAVAKIILAQMLHNLVKISITLDBENYLEKGNALHELDIRNAVLEOLT 882
 DB 872 QTLVLMNCIDITSDCCDITLKLOEKSSILCIDLGLNHIGVGMKFLCEALRK--PLCNLR 929
 QY 883 ALMLPWGCDVYG--SLSSLKHLLEVPOLYKGLKNWRITDIEIHLGAFPKKNPL 936
 DB 930 CLWL-WGCSIPFSCEDICSLSN-QSLVTLIDL-----GQNP 965

RESULT 7

Q9RLM5 PRELIMINARY; PRT: 1111 AA.
 ID Q9RLM5
 AC Q9RLM5
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE OOPASM SPECIFIC PROTEIN.
 GN MATER OR OPT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIH/SWISS;
 RA Tong Z.-B., Nelson L.M.;
 RT "A mouse gene encoding an oocyte antigen associated with autoimmune
 premature ovarian failure.";
 Endocrinology 0:0-0(1999).
 DR EMBL; AF074018; AAB51762.1; -.
 DR HSSP; P10775; 28NH.
 DR MGD; MG1:1345193; Mater.
 DR InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam; PF00560; LRR_2.
 DR PRINTS; PR00364; DISEASERLIST.
 DR SMART; SM00370; LRR_7.
 SQ SEQUENCE 1111 AA; 125501 MW; 4016A5D67A1C01F4 CRC64;

Query Match 4.6%; Score 242; DB 11; Length 1111;
 Best Local Similarity 20.0%; Pred. No. 1.1e-07;
 Matches 210; Conservative 151; Mismatches 395; Indels 296; Gaps 46;

QY 143 DOHHNRVQLTNGLLQLOS-----PCIIEGSGKSTLQRIAMWGSGKCKALTKF 197
 DB 166 DHHSPKMLSLDAFKYQKTFQPHITLLGKRGVGSALARSIVGMAQG--KLRQKM 223
 QY 198 KVFVFL-----RLSRAOGLEFETLCDQLLDIPGTRKOTFAMMLKLRLQRYVLLD 248
 DB 224 SFVIFSVREIKWTKSSLAQ--LIAKRCPSWDLVTKIMSQP-----ERLLFYID 272
 QY 249 GYN-----EKRPNCPKEL--EALIKENRFRNMVYVTTTECLRHINQFG 291
 DB 273 GLDDMDVYLQHDMDTLSDMKWDQEPYITLWVSLRKALLPSPFITTRNRTGLEKSM- 331
 QY 292 ALTAEVGMDTDSQAQLREVILKELAGLLQIOKSRCLNLT----- 334
 DB 332 -----VSPLYI--LVGCLASRSQVLLENISNESRIOVFSLIENH 373
 QY 335 -----MKTPLFVVTIC-AIQGES-----EFHSHQTQTLFHT--FYDLLIQKNHKKHKG 381

Db 374 QLFDCQAPSVCSLVECALQLOKLGKRCITLPCQTLTGIVATLVPHQTLK--RPSQAL 431
 Qy 382 AASDFIRSLDHGCDLALLEGVFSHK---FDFELQDVSSVNEVDLLTTGL-LCKYTAQRFKP 437
 Db 432 SOEEOITIVGIC--MMAEGVWTRSVFYDDDLKNYSIKESIELAFHNNILLQVGHNSQ 430
 Qy 438 KYVFFKRSFOEYTAGRR--LSSLTSHPEEYTKNGYLOKMSISDITSTYSSL-- 431
 Db 491 CYVEFSLSLQDFFAALYYVEGL-----EEMNHCFEINQSRISMEVKRTDTRLLGKM 544
 Qy 492 RYVCGSVEATRAVMKHLAAVYOHGCLGLSTAKRPLMRQESLOSQKNTQELKAINI 551
 Db 545 RLFLFGL---MKNKDLTKLEVLFEFVPIPTVE-OKLOHWVSLAQOVNGTSPMDTLDF-- 558
 Qy 552 NSFEEGHIHYOESTSKSALSOEFAFQGSXYINSQ---NIPDYLFDFEHLPCNCASA 608
 Db 599 ----YCLFESQDEEFVGAALKR-----FQEWLLINQKMDLKVSYC-----LKHQON- 642
 Qy 609 LDFIKLDFYGGAMASWEKAADTGGIHMEAPETIYIPRAVSLFPFNKKOEFRTLEVTLRD 668
 Db 643 LKAIRVDIRD--LISVDNTLELCPVYTVQF-----TQCKPRLMEWNGNFCVGLSLRN 693
 Qy 669 FSKLNQD-----ITYLKTISSATSLR-L 692
 Db 694 LKELDLDGDSLISORAKITLCELRNOSCRITQKTFKSAEYVSGIKHLKILFSNONLKYL 753
 Qy 693 QIKRCAGVAGSLSVLSTCK---NIYSLMVEASPLTIEDERHIT---SVTNLKTLSIH 744
 Db 754 NLGNTPMKDDMDKLACEALKHPKCSVETLRDLSCGLITIGEMISTLLISTTRKLS-- 811
 Qy 745 DLDONRLPGGLTSLGNL-----KMLTKLIMD 771
 Db 812 -LAKNRVGSMSITSLNASSSMCLQKLTLLDNCGLTPASCHLLVSALFSONMLTHLCLS 870
 Qy 772 NIKMNEDEATKLKLGKLN---LKKMCLFH-----LTHLS-----D 803
 Db 871 NNSLGTBGOVQLOFLNRPICALRLRLNHCNIVDAGYGLAMRLANNTRKLTHLSLTMP 930
 Qy 804 IGEG-MDYIKSLSEPCDEEIQVSCSLANAVKILAEONLHNLVLTSLDSENYLEK 862
 Db 931 VGDGAMKLCEALKKEPTCYQLELELVQCTQNCCEDLACMITTKHKLKSLDGNNAIGD 990
 Qy 863 DQNEALHELDRKNNLEQLTALMLPMGCDVQGSLSILKLEVPOLVKLGKNNRLTDT 922
 Db 991 KGIYTL-----CEGLKSSSSLR-----RLGLACKLTSN 1020
 Qy 923 EIRILGAFKPNLKNFOQLNLAGNRVSDGMIAFMGVFE-NIKOLVFPDFSTKEFLDP 981
 Db 1021 CCBALSLAISCNP--HNSLNLVKNDESTSGMLKLSAFQCPVSNLGIITLMOEYI--- 1075
 982 ALVRLSOLVLSKTLFLQEARLV--GMQFDDDD 1011
 1076 ARVRR---QLEVEYFVPHVYIDGDWYASDED 1104
 RESULT 8
 Q9JLR2 PRELIMINARY; PRT; 1111 AA.
 AC Q9JLR2;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MATERNAL-ANTIGEN-THAT-EMBRYOS-REQUIRE PROTEIN.
 GN MATER.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10990;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV; TISSUE=OOCYTE;
 RX MEDLINE=20222768; PubMed=10754103;
 RA Tong Z.B., Nelson L.K., Dean J.;

RT "water encodes a maternal protein in mice with a leucine-rich repeat
 RT domain homologous to porcine ribonuclease inhibitor."
 RL Mamm. Genome 11:281-287(2000).
 DR EMBL; AF143573; AAF64393.1; -.
 DR EMBL; AF143559; AAF64393.1; JOINED.
 DR EMBL; AF143560; AAF64393.1; JOINED.
 DR EMBL; AF143561; AAF64393.1; JOINED.
 DR EMBL; AF143562; AAF64393.1; JOINED.
 DR EMBL; AF143563; AAF64393.1; JOINED.
 DR EMBL; AF143564; AAF64393.1; JOINED.
 DR EMBL; AF143565; AAF64393.1; JOINED.
 DR EMBL; AF143566; AAF64393.1; JOINED.
 DR EMBL; AF143567; AAF64393.1; JOINED.
 DR EMBL; AF143568; AAF64393.1; JOINED.
 DR EMBL; AF143569; AAF64393.1; JOINED.
 DR EMBL; AF143570; AAF64393.1; JOINED.
 DR EMBL; AF143571; AAF64393.1; JOINED.
 DR EMBL; AF143572; AAF64393.1; JOINED.
 DR MGD; MGI:1345193; Mater.
 DR InterPro; IPR000767; Disease_resist.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR003592; LRR_out.
 DR Pfam; PF00560; LRR; 2.
 DR PRINTS; PR00364; DISSEASERST.
 DR SMART; SM00370; LRR; 7.
 DR SEQUENCE 1111 AA; 125544 MW; B07F9DF1B18DD855 CRC64;

Query Match 4.6%; Score 241; DB 11; Length 1111;

Best local Similarity 19.5%; Pred. No. 1,3e-07;
Matches 204; Conservative 158; Mismatches 399; Indels 286; Gaps 44;

Qy 143 DQNHREQLTGLGLQALQS-----PCIEGSGKSTLLQRIAMLSGSKALTRF 197
 Db 166 DLHYDEPEKKLLSDAKRPQKTFQPHITLHGRVGSKALASIVLGNAGS--KLFOKM 223
 Qy 198 KEVFFRLSLR---AAGLFETLQDLDIPGIRKQTFAMLLKRLQRYFLIDGYN-- 251
 Db 224 SFVIFFSVREIKWTEKSSLAOLIAKCPDSMDPVTK-----IMSOPERLLFYIDLDIM 277
 Qy 252 -----EFKPNCPRI--EALIKENHFRKNVYVTTTCTCLRHIRQFALTA 296
 Db 278 DSVLQHDMDTLSDMKDDEQPIYILMSLRKALLPQSFLLITTRNGLEKLSM----- 331
 Qy 297 VGDMDTSDAQLIREVLVLELAEGLLQIQSKRLNL----- 334
 Db 332 -----VVSPLYL--LVGGLASRRSQVLLENISNESDRIGQVSHLENHQLFDQ 378
 Qy 335 MKTPFLVVITC-AIQNGES-----EFHSHTQTLFRT--FYDLLQKNNKHKGVAASPF 386
 Db 379 CQAPSVCSLVCEALQLOKLGKRCITLPCQTLTGIVATLVPHQTLK--RPSQALSQEQ 436
 Qy 387 IRSLDHCGDLALLEGVFSHK---FDFELQDVSSVNEVDLLTTGL-LCKYTAQRFKPKFF 442
 Db 437 ITLVGIC--MMAEGVWTRSVFYDDDLKNYSIKESIELAFHNNILLQVGHNSQCYVS 495
 Qy 443 HKSFOEYTAGRR--LSSLTSHPEEYTKNGYLOKMSISDITSTYSSL-----RYTG 496
 Db 496 HSLDQEFALYYVEGL-----EEMNHCFEINQSRISMEVKRTRDTRLLGKMKFLFG 549
 Qy 497 SSVEATRAVMKHLAAVYOHGCLGLSTAKRPLMRQESLOSQKNTQELKAININSFVE 556
 Db 550 L--MKNKDLTKLEVLFEFVPIPTVE-OKLOHWVSLAQOVNGTSPMDTLDF-----Y 599
 Qy 557 CGIHLYOESTSKSALSOEFAFQGSXYINSQ---NIPDYLFDFEHLPCNCASALDFIK 613
 Db 600 CLFESQDEEFVGAALKR-----FQEWLLINQKMDLKVSYC-----LKHQON-LKAIR 647
 Qy 614 LDFYGGAMASWEKAADTGGIHMEAPETIYIPRAVSLFPFNKKOEFRTLEVTLRDSKIN 673
 Db 648 VLRD--LISVDNTLELCPVYTVQF-----TQCKPRLMEWNGNFCVGLSLRNKELD 698
 Qy 674 KQD-----ITYLKTISSATSLR-LQIKRC 697

DB 699 LGDSILSOPAMKILCELELNOSCRIOKLPKSAEVVSGLKHLMLKLFSSNQNLKYLNGMT 758
QY 698 AGVAGSLSLVLTCK-----NIVSIWEASPLTIEDERHIT-----SYTNKJTSIHDLQNO 749
DB 759 PKMDDMKLACALPKKCSVETLRIDSCETLTIGYEMISTILLISTRLKLS---LARN 815
QY 750 RLPGGLTDSIAGNL-----KNLTKLIMDNKKN 776
DB 816 RGVGVSMSISGNALSSMKCLLKLINDGCLTPASCHLLVSALFSSQNTLTHCLSNISG 875
QY 777 EDDAIKLAGLKN---LKKMLCFH-----LTHLS---DIGEG- 807
DB 876 TGVGOOLCOFLRNPECALQRLILNHCNIYDVAYGFLAMRLANNKTLHLSTLMNPYGDA 935
QY 808 MVIYKSLSSPECDLEIOLVSCCSANAVKLLAONLHNLVLSLIDLESENLEKDNNA 867
QY 936 MKLCEALKEPTCYLOELVDCQLTQNCEDLACMITTTKQLSLDIGNMLGKGVIT 995
DB 868 LHELIDRMNVLEQLTALMLPMGCDVQGSLSLKLHLEVPOLVKLGKRWRLTDEIRIL 927
QY 996 LCEGLKQNN-----SSLRRLGCGCKLTSNCCAL 1025
DB 928 GAFEGKNPLKNFOQLNAGNRVSSDGMALFMGVFE-NLKQVFPFSTKEPLPDALYRK 986
DB 1026 SLAISGNP--HNSLNLVKNDEFTSGMLKCSAFQCPVSNLGLIGLMQKEY--ARVRR 1080
QY 987 LSOVLSKLTFLQDEARLV--GNQFDDDD 1011
DB 1081 ---OLEEVEFKPHVYIDGMYASDED 1104

RESULT 9
09HC29 PRELIMINARY: PRT: 1040 AA.
AC 09HC29;
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE NOD2 PROTEIN.
CN NOD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
Ogura Y., Inohara N., Benito A., Chen F.F., Yamaoka S., Nunez G.;
"Nod2, a Nod1/Apa1 family member that is restricted to monocytes and
RT activates NF-kappaB";
RT J. Biol. Chem. 0:0-0(2001).
DR EMBL: AF178930: AAC3677.1:--
DR InterPro: IPR001687: ATP_GTP_A.
DR InterPro: IPR001315: CARD.
DR InterPro: IPR001611: LRR.
DR InterPro: IPR003592: LRR_out.
DR InterPro: IPR003590: LRR_RnInh.
DR Pfam: PF00560: LRR_2.
DR SMART: SM00114: CARD_1.
DR SMART: SM00370: LRR_5.
DR SMART: SM00368: LRR_RI_7.
DR PROSITE: PS50209: CARD_1.
SQ SEQUENCE 1040 AA; 115282 MW; 0037592D96D7DDFF CRC64;

Query Match 4.4%; Score 234.5; DB 4; Length 1040;
Best Local Similarity 20.1%; Pred. No. 3.2e-07;
Matches 214; Conservative 157; Mismatches 384; Indels 311; Gaps 42;

QY 4 IKDNRALIORBMGVYIKOITDGLVWVNLREVNIIICEKEV-----QDAAGIITHM 57
DB 134 IOSHRPAIVRLHSHVENKL--DL-AW--ERGFVSQYECDEIRLPIFTPSQARRRLDL 187

QY 58 ILKGSSECNLFKSLKLENNYPLFQDL-----NGOSLFHOTSEGD----- 97
DB 188 ATVKANGLAALFLLQIVQELPYLALPLEAATCKKTMALRTVSAQSFSTYDAEFLC 247
QY 98 LDDLAD-----IKDLYHTPSFLNFYPLGDDIDLIIFNLKSTF 134
DB 248 LEDIYENLEVMADVGMAGPQKSPATLGLELESTGHLN-----DDADTV----- 295
QY 135 TEPVLMRKDQHHRVLEQLTNGLLALQALSPCLIEESGKSTLLQRIAMMGSKCAL 194
DB 296 -----LVGEASGKSTLLQRIHLMAAG--ODF 322
QY 195 TKKFVFFLRLSRAO-----GGLFETLDDQLDIPGTRKOTFMAMLLKLKQRYLF 245
DB 323 QEFLEVFPPSCQLOCMKPLSVRTLLEHCCW-----DVGQDITQLLDHPDRVLL 376
QY 246 LLDGYNEFK-----PONCEIEAL-----IKENHFKN--VVIYTTTECLRHRQ 289
DB 377 TDEGDFEEFKFPTDREBRHCSPTDPTSVQTLLENLQGNLKKARKRVYNSRPAVSAFLRK 436
QY 290 FGALTAEVGDMTEDSAQALIREVLKE--LAEGLLIQIQKSRCLRLMTPLF--VITCA 346
DB 437 YIRTEFNKGFSEOGIEIYLKRHHEPGVADRILRLQETSLMGLCHLPVFSMWVSKCH 496
QY 347 IOMGESEFHSHTQTLTFHTFYDLIQKNKH-----KGAAS--DFIRSLDHGCD 395
DB 497 QELLQEGGSPKTTTDM-----LLLOHFLHATPPDSASQGLSRLGRPLTLHGR 552
QY 396 LALGEG-----VSHKPFDELQDVSSVNEVDLLTGLCKYTAQRFKRYFKHKSFOEY 449
DB 553 LALMGLGMCYVFSQAQ-----QLQAAQVSPDDISLGLFVRAKGVPGSTAPLEFLHTTFQCF 609
QY 450 TAGRLSLTLTSHPEEYTK-----GNGLOKMYNISDI-----TSYSSILRTIC 495
DB 610 FAAYFLA--LSADVPPALLRHLNLCGRPGNSPMALLPTMCIQASEGKSSVAALLQAAE 667
QY 496 GSSVETRAVMMKHLAAVYOHGCLGLSLAKRPLMROESLOSAYKNTFEOEILKAININSFV 555
DB 668 PHNLDTTAFLAGLSLREHWGLLACQTSERALLRQA----- 705
QY 556 ECGIHLVQESTSKSALSOFEAFFOGKSLYINSQNIPLYLDFEFHLPNCASALDFIKLD 615
DB 706 -CARMLARSLKHHHSIPPAARGEAKSVHAMPG-----FIWLI 743
QY 616 FYGAMASWEKAEDTGGIHMEARETYI--PSAAVSLFFEMKQEPRTLEVTLRDFSKL 672
DB 744 RSLVMEERLARKAARGLVNGHLKLTFCVSGPTCCALAFVLQHLRRVALQD----- 797
QY 673 NKODITTYLGIFFSSATSLRLQIKRCAGVAGSLVLTCKNIY-----SLAYEASPLTIED 728
DB 798 ---DYSVGDICVE-----QLLPCLGV-----CKALYLRDNNISDRGICKLIEC 838
QY 729 ERHITSVTNLKTLSTLHDLONLRLPGGLTDSIGNL-----KNLTKLIMDNKKNEDAIKLA 784
DB 839 ALHCEQLOKLA-----LFNNKLTGCAHSMAKILACQNFALRLGNNYITAAQAYLA 892
QY 795 EGIKLKMKKICLF-----HLTHLDIGCGMRYIKKSLSSPECD 821
DB 893 ECLRGNTSLDQFLGFWGNRVGDEGAQALAEALDQHSLEMLSVGN--TGSVGAQALA 948
QY 822 L---EELIVSCCLSAN-----AVKLLAONLHNLVLSLIDLESEVLEKDEGNALHEHID 873
DB 949 LMLAKNVNLEELCELENNILODEGVCSLADGLKKNSSKILKSNNCITTYLGAENLQALE 1008
QY 874 RKNVLEQLTALMLPMGCDVQGSLSLKLHLEVPOLVKLGKRWRL 919
DB 1009 RNDLTILEV-----W---LRGNTFS---LKGCRDFTL 1038

RESULT 10
09HAV8 PRELIMINARY: PRT: 1429 AA.
AC 09HAV8;

DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE CASPASE RECRUITMENT DOMAIN PROTEIN 7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RA BERTIN J., DISTEFANO P.S.;
 RT "The PYRIN Domain: A Novel Motif found in Apoptosis and Inflammation
 RT Proteins";
 RL Cell Death Differ. 0:0-0(2000).
 RP
 RP SEQUENCE FROM N.A.
 RA BERTIN J.;
 RL Submitted (AUG--2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF298548; AAC15254.1; -;
 DR InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003590; LRR_Rninh.
 Pfam: PF00560; LRR_1.
 PRINTS: PR00364; DISEASESISIT.
 SMART: SM00370; LRR; 5.
 SMART: SM00368; LRR_R1; 5.
 SQ SEQUENCE 1429 AA; 160880 MW; 90F809AB838B7F7 CRC64;

Query Match 4.18; Score 217; DB 4; Length 1429;
 Best Local Similarity 20.9%; Pred. No. 7.8e-05;
 Matches 185; Conservative 97; Mismatches 289; Indels 314; Gaps 36;

QY 142 KDQHHNREVDLTNGLLQALOSP--CIIEESGSGKSTLLQRIAMLGSGK----- 190
 DB 305 EENRGHILIEDLFGPELDQPEPRIVILQGAIGKSTLARQKAEANGROQLGDRFQHV 364
 QY 191 ---CAALTKEFVFYFLRLSRAOGLFETLCDLDLIPGTRKOTFAMLLKRLQRYVFL 246
 DB 365 FYESCRLAEKSKVYSLAELGKDG-----TATPAPIRQ-----ILSRPERLLFI 408
 QY 247 LDGNEKRPONCEIELEIKENHFKNMVITTTTECLRIHQFALTAEGVDMTEDSAQ 306
 DB 409 LDGVDL-----PGWVLOEPSSSELCILHWSQ-----PQPAD 437
 QY 307 ALIREVLIKE-LAEGLLLOQSKRLNKKTPLEFVVITCAIQMGSEFHSHTQTLFHT 365
 DB 438 ALLGSLGKTIPLASPLIARTALONLI--PSLEQARWVEVLGSE--SSKREYFRRY 493
 QY 366 FYD-----LLIQNKKHKKGVAAADFIRSL-- 390
 DB 494 FTDEQRAIRAFRLVYSNKKELMALCLVPWVSMIACTCLMOQMKRREKLTLSKTTTCLH 553
 QY 391 -----DHCGDLAEGVFSHKFPELQDVSSVND-----VLTTGLLCK 429
 DB 554 YLAQALQAPLGLQRLDLC-SLAEGIMORKTLLFSPDLKRHGLDGAIIITFLMKGLIOE 612
 QY 430 YTAQRFKPKKFFPKSROETAGRLSLSLTSHPEEVTNGCYLQAMVSISSDTSYSS 489
 DB 613 HP---PLLSFTHLCQEFPA--MSYVL-----EDEKGRG----- 644
 QY 490 LLRYTGSSEVATRAVKKHLAAYVQHCLGLSLAK--RPLMROESLQSVKNTTEQELK 547
 DB 645 -KHSNCLIDEKT-----LEAVGIHG-LFGASTFRFLGLLSDGSRKEMENIFHCRLSQ 696
 QY 548 AININSVEEC-----GIHLVQESTSKALSQEFPAFGKSLYINSGN---IPD 593
 DB 697 GRNIMQWVPSIQLDLPHSLESLHCLYETRNKFTLLQVM-AHFEEMQCMVETDMLLCT 755
 QY 594 YLDFEFHLNFCASALDFIKLDFYGGAMASWEKAAEDTGGIHMEAEVATIPSAVSLF- 652
 DB 756 FCIKFSNHRV-----KLQLEIG-----RQHSSTSPSKVAV-LFR 788

QY 653 -----FNNKQEFRTLEVTLRDESKLNKODITYLGKIFSSATSLRLQIKRCAVAGSL 706
 DB 789 WVPYDAVWQWLLFSVLKVT-RNKELD-----LSQNSLSHSA 824
 QY 707 VLTCKNITYSLWVASEPLTIEDERHITSYTNLKTLSIHDLOQRDLP-GLTDSLGNKLN 765
 DB 825 VKSLCK-----TLRRPRCL-----LETURLAGCGLT----- 850
 QY 766 TKLIMDNKMNEDAIKLAELGKMKMLFHLTHLSIDGEGMDYIKSLSPDCLEET 825
 DB 851 -----AECKDLATGLRANOTLTLELDSFNVLMAGAKHLQRLRQPSCKLQRL 899
 QY 826 QLVSCCLSANAVKILAONLHVLKSLTLDSENYLEKDGNEALHELIDRANVLEQTLAM 885
 DB 900 QLVSCGLTSDCCODLASVLSASPSLKELDLQONNLDQVGRLLCE----- 944
 QY 886 LPMGCDVQSGSLSLKHLEVPOLVIGLKNRILTD---TEIRIL 927
 DB 945 -----GLRH--PACKLIRLGLDOTPLDSEMROELRAL 974

RESULT 11

QY 090FT4 PRELIMINARY; PRT; 1192 AA.
 AC 090FT4;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE HYPOTHETICAL 134.6 KDA PROTEIN (FRAGMENT).
 GN DKEP258601822.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RA TISSUE-UTERUS;
 RC Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL117470; CAB55945.1; -;
 DR InterPro: IPR000767; Disease_resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 2.
 DR PRINTS: PR00364; DISEASESISIT.
 DR SMART: SM00370; LRR; 5.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 1192 AA; 134610 MW; D28D00457C36229A CRC64;

Query Match 4.0%; Score 212; DB 4; Length 1192;
 Best Local Similarity 20.8%; Pred. No. 1.3e-05;
 Matches 184; Conservative 97; Mismatches 290; Indels 314; Gaps 36;

QY 142 KDQHHNREVDLTNGLLQALOSP--CIIEESGSGKSTLLQRIAMLGSGK----- 190
 DB 24 EENRGHILIEDLFGPELDQPEPRIVILQGAIGKSTLARQKAEANGROQLGDRFQHV 83
 QY 191 ---CAALTKEFVFYFLRLSRAOGLFETLCDLDLIPGTRKOTFAMLLKRLQRYVFL 246
 DB 84 FYESCRLAEKSKVYSLAELGKDG-----TATPAPIRQ-----ILSRPERLLFI 127
 QY 247 LDGNEKRPONCEIELEIKENHFKNMVITTTTECLRIHQFALTAEGVDMTEDSAQ 306
 DB 128 LDGVDL-----PGWVLOEPSSSELCILHWSQ-----PQPAD 156
 QY 307 ALIREVLIKE-LAEGLLLOQSKRLNKKTPLEFVVITCAIQMGSEFHSHTQTLFHT 365
 DB 157 ALLGSLGKTIPLASPLIARTALONLI--PSLEQARWVEVLGSE--SSKREYFRRY 212
 QY 366 FYD-----LLIQNKKHKKGVAAADFIRSL-- 390

```

DB 213 FTDEROAIARFLVKSNEKELMALCLVPWVSWLACTCLMOQMKREKLTSTTTTLCIH 272
OY 391 -----DHGDLALGVESHKRFDFELQDVSSVNE-----VLTITGLICK 429
DB 273 YLAQALQAPLGPOLRDLCL-SLAAGIWMKTLFSPDDLARKHGLGAIISTFLKGILOE 331
OY 430 YTAQRFKRYKFFKSFQRYTAGRLSLTSHPEEYTKNGYLQKWSISDITSTYSS 489
DB 332 HP---IPLSYFIHLCFQEPFAA--MSYVL-----EDEKGRG----- 363
OY 490 LIRYCGSSVEATRAVMKHLAAVYOHGCLLSIAK--RPLMRQESLOSVMKNTPEQELK 547
DB 364 -KHSNCIIDELEKT-----LEAYGIHG--LFGASTTRFLGLISDGEREMENIFHCRLSQ 415
OY 548 AININSFEVC-----GIHLQESTKSALSOFEAFQOKSLYINSGN---IPD 593
DB 416 GRNLMQWVPSIQLLQPHSLSEIHLCLYETRNKTFPLQVW-AHFEEWMCVETDMELVCT 474
OY 594 YLDFEFELPNCASALDFIKLDFYGGAMASWEKAEDTGTGIMBEAPETIYPSRAVSIF- 652
DB 475 FCIKFSRHVK-----KIQLEIG-----RQHRSTWSPJWV--LFR 507
OY 653 -----FWMKQEFRTLEVTLTRDFSKLNODITYLQKIFSSATSLRLOIKRCAGVAGSLSL 706
DB 508 WVPYTDAYWQILFSLVKYT--RLKKELD-----LSGNSLSHSA 543
OY 707 VLSTCKNIYSLMVEASPLTIEDERHITSVTNLKTLISHDLQONRLPG--GLTDSLGNLKNL 765
DB 544 VKSLCK-----TLRRPCL-----LETLRLAAGGLT----- 569
OY 766 TKLIMDNKMNEDAIKLAEGKLNKMKCFHLTHLSDIGEGMDYIVKSLSEPCDLEI 825
DB 570 -----AEDCKDLAFGLRANQTLTELDLSFNVLTDAGAKHLCORLPSPCKLQRL 618
OY 826 QLVSCCLSANAVKLLAONLHNLVKLSILDSENYLEKDGNEALHEDRMVLEQLTALM 885
DB 619 QLVSGCLNSDCODLASVLSASPSLKELDLQONLDDVGVRLCE----- 663
OY 886 LPMGCDVQSGSLSLKHLHEEYQVLYKLGKIKMRLLD---TERIL 927
DB 664 -----GLRH--PACKLIRLIGLDTTSLDEMRQELRAL 693

RESULT 12
OY12E0 PRELIMINARY; PRT: 1429 AA.
OY12E0:
DB 01-NOV-1999 (TREMBLrel. 12, Created)
DB 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DB 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DB KIAA0926 PROTEIN (NALP1) (NAC-ALPHA SPLICED VARIANT).
DB KIAA0926 OR NAC.
DB Homo sapiens (Human).
DB Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DB Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
DB NCBI_TaxID=9606;
DB [1]
DB SEQUENCE FROM N.A.
DB TISSUE=BRIN;
DB MEDLINE=99246063; PubMed=10231032;
DB Naase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
DB Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
DB "Prediction of the coding sequences of unidentified human genes. XIII.
DB The complete sequences of 100 new cDNA clones from brain which code
DB for large proteins in vitro.";
DB DNA Res. 6:63-70(1999).
DB [2]
DB SEQUENCE FROM N.A.
DB Martinon F., Hofmann K., Tschopp J.;
DB "NALP1 a novel NACHT, LRR and PYD containing protein.";
DB Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DB [3]

```

```

RP SEQUENCE FROM N.A.
RA Chu Z.-L., Xie Z., Godzik A., Reed J.C.;
RT "NAC: an Apaf-1/Ced-4 family member regulates the cytochrome c pathway
RT for apoptosis.";
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
DB EMBL: AB023143; BAA76770.1; -
DB EMBL: AF310105; AAG30288.1; -
DB EMBL: AF229059; AAK00748.1; -
DB InterPro: IPR000767; Disease_resist.
DB InterPro: IPR001611; LRR.
DB InterPro: IPR003592; LRR_out.
DB Pfam: PF00560; LRR; 2.
DB PRINTS: PR00364; DISEASERIST.
DB SMART: SM00370; LRR; 5.
DB SEQUENCE 1429 AA; 160944 MW; 6C5CB8EFD2819435 CRC64;

Query Match 4.0%; Score 212; DB 4; Length 1429;
Best Local Similarity 20.8%; Pred. No. 1,7e-05;
Matches 184; Conservative 97; Mismatches 290; Indels 314; Gaps 36;

OY 142 KDHHHREVLTLNGLDALDSP--CITEGSGKSTLQRIAMWGSK----- 190
DB 305 EENRGHLIEIRDLFGPGDTPRIVILQGAAGIKSTLARQVKEWRCGLYGDPRQHV 364
OY 191 ---CKALTKRFVFLRLSRAQGLFETLQDLDIPGTRKQTFMMLKLRQVLEL 246
DB 365 FYFSRELAQSKVSLAELIGKG-----TATPARPQ-----ILSRPELLFI 408
OY 247 LDGYNEFRPQNCPEIEALIKENHFRKNMVIYTTTECLRHIRQFGALTAEGVDMTEDSAQ 306
DB 409 LDGVDE-----PGWVLDPESELCLHNSQ-----POPAD 437
OY 307 ALIREVLIKE-LAEGLLQIQSKRLNLMKTPYFVITCALIQWGESFPHSHQTTLFHT 365
DB 438 ALLGSLICKTLTPEASPLTARTALQNL--PSLEQARWVEVLGFSE--SSREYEFYR 493
OY 366 FVD-----LTIQKKKKKKGAASDFIRSL--- 390
DB 494 FTDEROAIARFLVKSNEKELMALCLVPWVSWLACTCLMOQMKREKLTSTTTTLCIH 553
OY 391 -----DHGDLALGVESHKRFDFELQDVSSVNE-----VLTITGLICK 429
DB 554 YLAQALQAPLGPOLRDLCL-SLAAGIWMKTLFSPDDLARKHGLGAIISTFLKGILOE 612
OY 430 YTAQRFKRYKFFKSFQRYTAGRLSLTSHPEEYTKNGYLQKWSISDITSTYSS 489
DB 613 HP---IPLSYFIHLCFQEPFAA--MSYVL-----EDEKGRG----- 644
OY 490 LIRYCGSSVEATRAVMKHLAAVYOHGCLLSIAK--RPLMRQESLOSVMKNTPEQELK 547
DB 645 -KHSNCIIDELEKT-----LEAYGIHG--LFGASTTRFLGLISDGEREMENIFHCRLSQ 696
OY 548 AININSFEVC-----GIHLQESTKSALSOFEAFQOKSLYINSGN---IPD 593
DB 697 GRNLMQWVPSIQLLQPHSLSEIHLCLYETRNKTFPLQVW-AHFEEWMCVETDMELVCT 755
OY 594 YLDFEFELPNCASALDFIKLDFYGGAMASWEKAEDTGTGIMBEAPETIYPSRAVSIF- 652
DB 756 FCIKFSRHVK-----KIQLEIG-----RQHRSTWSPJWV--LFR 788
OY 653 -----FWMKQEFRTLEVTLTRDFSKLNODITYLQKIFSSATSLRLOIKRCAGVAGSLSL 706
DB 789 WVPYTDAYWQILFSLVKYT--RLKKELD-----LSGNSLSHSA 824
OY 707 VLSTCKNIYSLMVEASPLTIEDERHITSVTNLKTLISHDLQONRLPG--GLTDSLGNLKNL 765
DB 825 VKSLCK-----TLRRPCL-----LETLRLAAGGLT----- 850
OY 766 TKLIMDNKMNEDAIKLAEGKLNKMKCFHLTHLSDIGEGMDYIVKSLSEPCDLEI 825
DB 851 -----AEDCKDLAFGLRANQTLTELDLSFNVLTDAGAKHLCORLPSPCKLQRL 899

```



```

Db 554 YLAQALQAPLGRDLC-SLAAGIWKTKLPSPDLRRKHGLDGAISTFLKMGILQ 612
QY 430 YTAQRFKRYKFFPKSPOEYTAGRLSSLTSHPEEYTKNGYLQKNVSLSDITSTSS 489
Db 613 HP---IPLSIFHLCROEFPAA--MSYVL-----EDEKGRG----- 644
QY 490 LLRYTCSSVEATRAVMKHLAAYOHGCLGLSLAK--RPLMRQESLQSVKNTTEOELK 547
Db 645 -KHSNCITIDEKT-----LEAYGIHG-LFGASTTRFLGLLSDGEGEMENIFHCRLSQ 696
QY 548 AININSFVEC-----GHLVQESTSKSALSQEFEEAFQOKSLYINSN---IPD 593
Db 697 GRNLMQWVPSLQDLLQPHSLSLHCLYETRNKTFLTQVM-AHFEEMGCYETDELLVCT 755
QY 594 YLDFEFELPNCASALDFIKDFYGGAMASWEKAAEDTGGHMEAPETYIPSRAYSLF- 652
Db 756 FCIFSRHVK-----KQILIEG-----ROHSTWSPYVW-LFR 788
QY 653 -----FNMKQEFRTLEVTLRDFSKLNKQDITYLGKIFSSATSLRLQIKRAGVAGSL 706
Db 789 WVPYDAVWQILFVSVLKYT-RNLKELD-----LSGNSLSHSA 824
QY 707 VLSCKKNYISLWVDSPLTIEDERHITSVNLKTLSHDLQONRPLG-GLTDSIGNLKNL 765
Db 825 VKSLCK-----TLRRPRL-----LETFLRAGCGLT----- 850
QY 766 TKLMDNKNNEEDAIKLAEGIKNLKMKCLFHLTHLSDIGGMDYIYKSLSSPCDLEI 825
Db 851 -----AEDCKDLAFGLRANQTLTELDLSFNVLTDGAKNHLQRLRQPSCKLQRL 899
QY 826 QLVSCSLANAVKILIAONLHNLVKLSILDSENYLERKDGNELHE 870
Db 900 QLVSCGLTSDCCQDLASVLSASPSLKELDLQONNLDDVGVALLCE 944

RESULT 15
Q9BZ28 PRELIMINARY; PRT; 1443 AA.
AC Q9BZ28:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NAC-DELTA SPLICER VARIANT.
GN NAC.
OS Homo sapiens (Human).
TX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chu Z.-L., Plo F., Xie Z., Godzik A., Reed J.C.;
RT "NAC: an Araf-1/Ced-4 family member regulates the cytochrome c pathway
for apoptosis."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF229062; AAK0751.1;
SQ SEQUENCE 1443 AA; 162237 MW; C30EB9BE9EC82FE96 CRC64;

```

```

Query Match 3.98; Score 207.5; DB 4; Length 1443;
Best Local Similarity 21.0%; Pred. No. 3.4e-05;
Matches 173; Conservative 91; Mismatches 280; Indels 281; Gaps 33;

```

```

QY 142 KDQHHNREQDTLGLQALQSP--CIIEGSGKSTLQRIAMINGSGK----- 190
Db 305 EENRGHLEITDLFGPLDQEPRIYILQGAAGIGKSTLARQKAWGSGQLYGDROHV 364
QY 191 -----CKALTKRFVFFPLISRAOGLEFETLCDQLDIPGIRKQTFMAMLLKRLQRYFL 246
Db 365 FYFSCRELAAQKVVSLAELGKDG-----TATPAPIRQ-----ILSPRELIFT 408
QY 247 LDGYNERRPONCPREIALIKENHRKKNVYTTTECLRHTRQFGALTAEVGDMTEDSAQ 306
Db 409 LDGVDE-----PCWVLQEPSSSELCLHWSQ-----POPAD 437

```

```

QY 307 ALIREVLIKE-LABGLLIQIQRSLRNLKMTPLFVYITCAIQMGSEFHSHTQTLFHT 365
Db 438 ALDGLSLCKTILPEASFLITARTALQMLI--PSLEQARWVEVLGFSE--SSREYFIRY 493
QY 366 FYD-----LLIQKNKHKHKGVAADFTIRSL--- 390
Db 494 FTDERQAIRARIVKSNKELMALCLVPVWSVLACTCLMQQKRRKREKLLTJTKTTTLC 553
QY 391 -----DHGDLALGVFHSKFEDELQDVSSVND-----VLTGGLCK 429
Db 554 YLAQALQAPLGRDLC-SLAAGIWKTKLPSPDLRRKHGLDGAISTFLKMGILQ 612
QY 430 YTAQRFKRYKFFPKSPOEYTAGRLSSLTSHPEEYTKNGYLQKNVSLSDITSTSS 489
Db 613 HP---IPLSIFHLCROEFPAA--MSYVL-----EDEKGRG----- 644
QY 490 LLRYTCSSVEATRAVMKHLAAYOHGCLGLSLAK--RPLMRQESLQSVKNTTEOELK 547
Db 645 -KHSNCITIDEKT-----LEAYGIHG-LFGASTTRFLGLLSDGEGEMENIFHCRLSQ 696
QY 548 AININSFVEC-----GHLVQESTSKSALSQEFEEAFQOKSLYINSN---IPD 593
Db 697 GRNLMQWVPSLQDLLQPHSLSLHCLYETRNKTFLTQVM-AHFEEMGCYETDELLVCT 755
QY 594 YLDFEFELPNCASALDFIKDFYGGAMASWEKAAEDTGGHMEAPETYIPSRAYSLF- 652
Db 756 FCIFSRHVK-----KQILIEG-----ROHSTWSPYVW-LFR 788
QY 653 -----FNMKQEFRTLEVTLRDFSKLNKQDITYLGKIFSSATSLRLQIKRAGVAGSL 706
Db 789 WVPYDAVWQILFVSVLKYT-RNLKELD-----LSGNSLSHSA 824
QY 707 VLSCKKNYISLWVDSPLTIEDERHITSVNLKTLSHDLQONRPLG-GLTDSIGNLKNL 765
Db 825 VKSLCK-----TLRRPRL-----LETFLRAGCGLT----- 850
QY 766 TKLMDNKNNEEDAIKLAEGIKNLKMKCLFHLTHLSDIGGMDYIYKSLSSPCDLEI 825
Db 851 -----AEDCKDLAFGLRANQTLTELDLSFNVLTDGAKNHLQRLRQPSCKLQRL 899
QY 826 QLVSCSLANAVKILIAONLHNLVKLSILDSENYLERKDGNELHE 870
Db 900 QLVSCGLTSDCCQDLASVLSASPSLKELDLQONNLDDVGVALLCE 944

```

Search completed: March 25, 2002, 10:59:51
 Job time: 246 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2002, 10:55:00 ; Search time 21.16 Seconds
(without alignments)
3686.330 Million cell updates/sec

Title: US-09-697-089-2

Perfect score: 5281

Sequence: 1 MNFIKDNSRALIQMGMTVI.....WQFDDDLVITGAFKLVTA 1024

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR-68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	632.5	12.0	1447	2 T42628	neuronal apoptosis
2	561.5	10.6	1232	2 A55478	neuronal apoptosis
3	212	4.0	1192	2 T17255	hypothetical prote
4	198.5	3.8	1590	2 B86398	protein T7N9_24 [1
5	187	3.5	1405	2 T04426	hypothetical prote
6	178.5	3.4	1031	2 G96663	hypothetical prote
7	176	3.3	862	1 PAD0A	alpha-actinin - sl
8	171.5	3.2	1167	2 T06146	disease resistance
9	170.5	3.2	1220	2 T06403	resistance complex
10	170	3.2	1066	2 T15864	hypothetical prote
11	169.5	3.2	1110	2 F84547	probable disease r
12	165	3.1	1112	2 T10504	disease resistance
13	164.5	3.1	1112	2 T06145	disease resistance
14	164	3.1	1016	2 T30553	disease resistance
15	164	3.1	1060	1 A40264	kinesin-related pr
16	163	3.1	853	2 T17461	disease resistance
17	162	3.1	1784	2 C96615	hypothetical prote
18	160.5	3.0	1133	2 E86308	hypothetical prote
19	160	3.0	1217	2 T51141	disease resistance
20	159	3.0	921	2 B86234	hypothetical prote
21	159	3.0	1120	2 B86479	hypothetical prote
22	159	3.0	1215	2 H84513	probable disease r
23	159	3.0	1217	2 T51140	disease resistance
24	158.5	3.0	568	2 F86291	hypothetical prote
25	158.5	3.0	766	2 T07039	hypothetical prote
26	153	2.9	845	2 T07039	Hc9-0 protein - t
27	153	2.9	919	2 T05746	hypothetical prote
28	153	2.9	1019	2 C96519	probable disease r
29	152.5	2.9	526	2 C84552	hypothetical prote

30	152.5	2.9	2429	1 SJHUA	spectrin alpha cha
31	151.5	2.9	1305	2 T18548	flax rust resistan
32	151	2.9	855	2 T17460	disease resistance
33	151	2.9	4767	2 T31345	hypothetical prote
34	150.5	2.8	1234	2 T14515	hypothetical prote
35	150	2.8	1269	2 F84730	probable myosin he
36	150	2.8	1277	2 S54451	hypothetical prote
37	150	2.8	2269	2 T28677	hypothetical prote
38	148.5	2.8	1447	2 F82909	hypothetical prote
39	148	2.8	1830	2 E82909	conserved hypothet
40	148	2.8	2167	2 S50658	bud emergence prot
41	147.5	2.8	957	2 E84547	probable disease r
42	147.5	2.8	1011	2 C84524	probable disease r
43	146.5	2.8	2401	2 T28676	hypothetical prote
44	146.5	2.8	2493	2 S45734	probable membrane
45	146	2.8	2253	2 T30336	nuclear/mitotic ap

ALIGNMENTS

RESULT 1	T42628	neuronal apoptosis inhibitory protein 2 - mouse
C:Species:	Mus musculus (house mouse)	
C:Date:	11-Jan-2000	#sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession:	T42628	
R:Taraghi, Z.; Diez, E.; Gros, P.; Mackenzie, A.		
Mamm. Genome 10, 761-763, 1999		
A:Title:	CDNA cloning and the 5' genomic organization of Naip2, a candidate gene for m	
A:Reference number:	222179; MUID:99315542	
A:Accession:	T42628	
A>Status:	preliminary; translated from GB/EMBL/DBJ	
A:Molecule type:	mRNA	
A:Residues:	1-1447 <VAR>	
A:Cross-references:	EMBL:AF102871; NID:g3860228; PID:g3860229; PIDN:AAC73002.1	
A:Genetics:		
A:Gene:	Naip2	
Query Match	12.0%; Score 632.5; DB 2; Length 1447;	
Best Local Similarity	24.2%; Pred. No. 3.1e-29;	
Matches	268; Conservative 195; Mismatches 423; Indels 221; Gaps 44;	
OY	38 VNIICEVEVDAAAGITIMILKSGESCNFLKSLKEMNIPFLFDLNGQSLFHQTSBGD 97	
DB	404 VSVLCRDQ-DHSEAG-----RCGASSGYLPS-----TDL-GQSEACWLGQA- 444	
OY	98 LDDLAQDLKDLVHTPSPLNF-----YPLGEDIDIFNLKSTFTEPVLARKDQH 145	
DB	445 -RSLSQRLDVTYTKATFERHMLPEVYSSLGTDHLSCDVSTI---SKHISQV----- 493	
OY	146 HHRVQTLTNGLLQALQSPCIIIEGSGKSTLQRIAMLMGSGKALTFKFEFFRL 205	
DB	494 ---GQSLTIPFVSNLNVCMCEGASGKTFELKRIAFVMSGCCPLINRQVLVFTSL 550	
OY	206 SRAQGC--LFTFLCDQLDIPETIRKQTFMAMLKLQRQVPLFDGVYEFK--PONCEI 261	
DB	551 SSIPGQELAKICQALDLAGGCCISEVCSIIQQLQVFLFDVYSGLASLPQ---AL 607	
OY	262 EALIKENRFRKMNIVTTTECLRIHROFALTAAGVGMTEDSAQALREVLIKRLAGL 321	
DB	608 HTLITKNLSTRCLIAVHTNKVKGIRYLDTSLEIKFPPRYIVSVLKLFSHD----- 662	
OY	322 LLAQDK-----SRCLNLMKTPLEFVITCAIQMGSEFHSHTQTLFHTFYDLLQKN 374	
DB	663 IMRVAKFPIVNGFHELGIGIKHTPLFVAAYCQDMFKNPSDQPFQDVAFKAYMGL----- 718	
OY	375 KHKHGVASDFIRSLDCGDLALEGVSHKFDPELDV--SSVNEVDLLTGLGLCKTA 432	
DB	719 SLKHKG-AAKPLQAVSSGQDALALGLFSSCFEFNSMDIAGVDEBELDTTCLMSKPTA 777	
OY	433 QRFKRYKFFPKHSFQEVYAGRRSLSLTSHPEEYTKNGVLQMKVYSLSDTITYSLSLR 492	

```

Db 778 QRRPYRRFGLPFOEFLAVALRLTELLSSDRDODGLGYLLROINSPLKAMSITYHFLK 837
QY 493 YTCG--SSVEATRAVMHAAVYQHGLGLSLKRPMLRQESLOSQYNTQEOELKATINI 551
Db 838 YVSHSSSKAAPVYVSHL-----LQVDEKESLENMSENEDYKMLHP 879
QY 552 NS--FVECGIHLVQ-----ESTSKSALSOEFAFOGKSLYI 586
Db 880 EALLTEICLGLMQLSPESFLISENLRLICLNFAHESNTVAACSPVLIQLGRRLDL 939
QY 587 NSGNIPDYLFDFEHLPCNSALDFIKLDYGGAMA-----SWEA----- 627
Db 940 KVLST-QY--FWDH--PETLLLSKIKISLGNMNVQRIDESLEKSEKVOPTIDQY 994
QY 628 -----AEDTGSIHMEAPETYIPSRVAVSLFENMK-----OEFTLEVTLRD 668
Db 995 AIAFQPINEYOKNLSEKKHAIKKEEDMKHQLPILNISTGY--WKLSPPRYIIPLEVOVTN 1052
QY 669 FSKLNKODITYYLGKIFSSATSLRLQIKRCAGVAGSLSLVSTCK-NIYSLMVEASPLTIE 727
Db 1053 TGPADQALLQVLMVEFSAQSIEFRISDSSGFLIESIRPALELSKASVTKCSMSRLSRE 1112
QY 728 DERHITSVTNLKILSHDLONQRLPGGLTSLGNLKNLTCLI----- 769
Db 1113 DOKILLTLPLQSLSLEVS--ETNQLPDLFNLHKLFLGLKELCVRLDSKRPVLSVLPGEFP 1170
QY 770 ----MDNIK---NEEPAIKLAEGKLNKKMKCLFHL--THLSIDGEMDYIVKLSSEPC 820
Db 1171 NLHMEELSTRSTESPLSKVLKIQNSPNIHFHLKCNLSNC-EPLMVLVASCCK--- 1226
QY 821 DLEIQLVSCCLSA-NAVKILIAQNLNLVNLKILSDL--SENVLEKDGNEALHELIDRNVL 878
Db 1227 -LRRIEFGRCFEMAFVNIILP---NFVFLKILNLRDQFPDKETSEKFAQALGSRLN 1281
QY 879 EQLTALMPLMGCDVQGLSSLSLKHLEVPOLVKIGLKNMKRLTPEITIL-GAFGKAPLK 937
Db 1282 EK---LFPVPGDGIHQVAKLIVRCQLPCLRLVLFVFAETLDDSVLEIARGATRG----- 1333
QY 938 NFOOL-UL---AGNRVSSDGLMFAFMGVFENLKOLFDFSTKEFLPD-----PALVRLS 988
Db 1334 GFQKLEMDLTLNKHITREBGRNFQVLDNLPNKNLDIS--RIIPFCIQIQAITYVALG 1391
QY 989 QVLSKITFLOEARLVGNQFDDDLSTV 1015
Db 1392 QCVSRILPSLRLGLMSVLDEEDIKVI 1418
ULT 2
3478
neutonal apoptosis inhibitory protein - human
N:Alternate names: NAIP
C:Species: Homo sapiens (man)
C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 02-Feb-2001
C:Accession: A55478
R:Roy, N.; Mahadevan, M.S.; McLean, M.; Shuttler, G.; Yareghli, Z.; Farahani, R.; Baird, S.
d, T.O.; de Jong, P.J.; Surch, L.; Ikeda, J.E.; Korneluk, R.G.; Mackenzie, A.
Cell 80, 167-178, 1995
A>Title: The gene for neuronal apoptosis inhibitory protein is partially deleted in ind
A:Reference number: A55478; MIMD:95112344
A:Accession: A55478
A:Molecule type: mRNA
A:Residues: 1-1232 <ROY>
A:Cross-references: GB:019251
C:Genetics:
A:Gene: GDB:SMA6; SMA
A:Cross-references: GDB:120378; OMIM:600354; OMIM:253300
A:Map position: 5q12.2-5q13
C:Keywords: apoptosis; ATP; glycoprotein; nucleotide binding; P-loop; transmembrane prot
F:479-496/Domain: transmembrane #status predicted <TMM2>
F:479-496/Domain: transmembrane #status predicted <TMM2>
F:476/Binding site: ATP (Lys) #status predicted

```

```

F:618, 632, 823, 923, 1035/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 10.6%; Score 561.5; DB 2; Length 1232;
Best Local Similarity 24.1%; Pred. No. 3.8e-25;
Matches 226; Conservative 155; Mismatches 341; Indels 215; Gaps 33;
QY 69 FLKSLKMNYPLEFQDLNGQ---SLFHQTSQGLDD----- 100
Db 344 FLQNMKS-SAEVTPRDLQSRGELCELLETTSNLEDSIAVGPIVEMAGQAOFOEAKN 402
QY 101 LAQDLADLYHTPSFLN-----YPLGEDDIINLKSTFEPYLMKKDHNHR 148
Db 403 LNEQLRAAYTSASFRRHMSLDDISSDLATDHLDCDLSI---ASKHISKPV----- 449
QY 149 VEOLITGLLQALQSPICIEGSGKSTLQRIAMLMGSGKCALTKFKVFFRLS-- 206
Db 450 QEPVLVLEPVGNLNSVWCVEGSGKTVLLKKAIFAMASGCCPLNRQVLVYLSST 509
QY 207 RAQGLFETLCDLDLIPGTRKQTFMAMLLKLQRYVFLLDGYNEF--KPOCPEIEAL 264
Db 510 RPEGGLASTICDQLLEKESVTEKCMRNIIQGLKNQVFLDDYKEICSIPO--VIGKL 566
QY 265 IKENHRKKNVYITTTTECLRHIRQSGALTAEVGDMTDSAOALIREVLKELA--EGLL 322
Db 567 IOKNHLSTRCLLIAVRNRAIRRYLETILIEIQAFPFYTVOLIRKLFSHMTRLKFM 626
QY 323 LQIQKSRLCLNMLKTPLEVATICAQMGSEPHSHQTQTFHFHYFDLQKNHKKHGA 382
Db 627 VYEGKQSLQKIQKTPFLVAICAHMFQYPPDSFDDVAVFKSYMELSLRNK-----A 680
QY 383 ASDFIR-SLDHCGDLAEGVSHKPFDELQDV--SSYNEVDLITGLCKYTAQREPKY 439
Db 681 TAEILKATVSSCGELALGFSCCFEENDDLAAGVDEDEDLMCLMSKFAQRLRPFY 740
QY 440 KPFHKSFOETAGRLSLTSHPEPEVTGNGYLQKMAVSIQITSYSSLLKRTGSSV 499
Db 741 RFLSPARQEFELAGRLIETLDSRQEHQDLGLYHLKQINSPPMTVASVYNNFLNV--SSL 788
QY 500 EATPA---VKKHIAAAYQHCGCLGLSLAKRPMRQESLOSQVKN-----TTBOEI 545
Db 799 PSTAGAKIYSHLHLVDN-----KELENISENDDVLKQNPETLSLOML 843
QY 546 LKAI---NINSFVECGIHL-----VOESTSKSALSOEFAFOGKSLYINSNPID 593
Db 844 LRGLMQICPOAYFESVSEHLLVIALKTAQOSNT--VAACSEFVLQFLOGRTLTGALNL-Q 901
QY 594 YLPD-----FPEHLPCN--ASALDFIKIDFYGG--AMASWE 625
Db 902 YFEDHPESLSLRSIHPSIRGNKTSPPRAHFSVLETCPDKSOVPTIDODVASAEPMNWE 961
QY 626 -----KAEDTGSIHMEAPETY-IPSRVAVSLFENMKOEFTLEVTL 666
Db 962 RNLAEREDNKSVMQMRKRSPLDSTGYKLSRQYKIP-----CLEVDV 1006
QY 667 RDESKLNKODITYYLGKIFSSATSLRLQIKRCAGVAGSLSLVSTCK-NIYSLMVEASPLT 725
Db 1007 NDIDVQCDMLEITMVFSAQSRIELHLNHSRGFIIESIRPALELSKASVTKCSMSRLS 1066
QY 726 IEDRHITSTYNTKLSIH---DLQNRPLPGGLTSLGNLKNLTCLMKNIM----- 775
Db 1067 AAEQELLTLPSLESLVSGSTIOSQDIFPN--LDKFLCKLELSVDLEGININFSVIPPE 1124
QY 776 -----NEEPAIKLAEGKLNKKMKCLFHLFHL-----SDIGGMDYIVKLS 816
Db 1125 FPNFHHMEKLLIQISAEDYSKLVASLPNFIISKLNLNLEBOQPPDETSKFAIYILGSL 1184
QY 817 SEPDLIEIQLVSCCLSANAVKILIAQNLNLVNLKILSDL 853
Db 1185 ----NEELLIPGTDGIYRAKALLIQCCQDLHCLRVL 1217
RESULT 3

```


[illegible]

RESULT 5

hypothetical protein T18B16.20 - Arabidopsis thaliana
 N:Alternate names: hypothetical protein F13C5.220
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 C:Accession: T04426; T05042
 R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
 submitted to the Protein Sequence Database, April 1998
 A:Reference number: Z15359
 A:Accession: T04426
 A:Molecule type: DNA
 A:Residues: 1-1405 <BEV>
 A:Cross-references: EMBL:AL021687
 A:Experimental source: cultivar Columbia; BAC clone T18B16
 R:Bevan, M.; Pohl, T.; Weizsaecker, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue
 submitted to the Protein Sequence Database, February 1998
 A:Reference number: Z15395
 A:Accession: T05042
 A:Molecule type: DNA

A:Residues: 1-1405 <BEM>
A:Cross-references: EMBL:AL021711
A:Experimental source: cultivar Columbia; BAC clone F13C5
C:Genetics:
A:Map position: 4
A:Introns: 893/3, 1154/2, 1183/3, 1192/3, 1198/2, 1214/3, 1251/1, 1282/2, 1327/1, 1381/2
A:Note: T18B16,20; F13C5,220

Query Match	3.5%;	Score 187;	DB 2;	Length 1405;
Best Local Similarity	20.1%;	Pred. No. 0.0056;		
Matches 201;	Conservative 151;	Mismatches 325;	Indels 322;	Gaps 51;

OY	163	SPCIIIEGSGKSTLIORIMIMSGSKALTFKFR--VFELIRSR--	OGGLFETLD	21	OY	27	SKTVLVBGAGIKTWLAKVES-----	ORVQEKVNLMLHNKKIEDEKSLYELIAA	78												
OY	219	QL-----	LDIP-GTIRKOTFMAMLIKLRQVLFELLDGYNFKPOMCP-	259	OY	219	QL-----	LDIP-GTIRKOTFMAMLIKLRQVLFELLDGYNFKPOMCP-	259												
OY	79	QLSTIIIEEGBEEDLDYPLESLEKIKREMKIKKKDKNDLLIDDEGSMTEDEYMOEL	138	OY	260	EIEALIKENHFKMMVYVTTTECLBRIHROGMLTAVGSMTEDSQALIRE-----	311	OY	139	NLODFLEKYSVK--	ILVTRDPE--	REEKE--	STTIKVGPLTKESJDLHDHEDLLTFS	192							
OY	312	-----	VDI-----	KEIABGLLIQIOKSRCLRNLMKPTPELVITCAIOMGESEFHSHTQ	359	OY	193	TSEDMPVLKRLCDNKEIKETPLMS-----	CLISKSGLPALIVLIKSLNISKMSAKO	247	OY	360	TTLEHTF-----	YDLIOKNNKHKHCVASDPIRSL	390						
OY	248	RKIKRELLSKSLDEAASAKNAIDRSRYNPVLQSLSTELKLPDETVKRYPIAC--	FWHIL	305	OY	391	D-----	HGCDL-----	ALEGVFS--	HKRFPELODVSVNEDVLLTGLLCKVTYAOERK	436	OY	306	DFKYSOGAYYRDLIVHMLEGYDPKYSVDAKVOBESHILMD--	FNNRGILKIQEDNMVY	364					
OY	437	PKYK-----	FFHKs-----	POETTAGRLSSLTSHREPEYTKGNCYLOKMSID	482	OY	365	PEFSMSNLLDODCGFFGRSSLGDRVYG-----	DKRRGLG--	KIILID	408	OY	483	ITSYKS-----	LEWYTCGSSVEATRAWKHIAVYQ--	HGCLIGLSIAK	525				
OY	409	MIQIOKSKNMITIILNSGNLNRREVGHKFFKPEMDLEVVLVFEETPHIELVLSLSKIK	468	OY	526	RPLMRQESLOSVAKNTTQEOIILKAININSFVECGIHLVQESTSKSAISOEFAPFOCKSLY	585	OY	469	K-----	LRVLVIRCDLDIDNIDIKLSGLO--	GLHVEVSGASSLV-----	505	OY	586	INSGNIPYLRDFPEHLRNPCASALDFIKLDFYGGAMSWKKAEDNGI-----	HW	636			
OY	506	-----	NIPD-----	DFKNNMTOLOS-----	LNSGLAISSPSTIEKLSMLKCFILRHCSL	552	OY	637	BEAPETYIPRSRAVSL-----	FFNNKQOERTEVLTLRDFSKLNKODITYLGIKFS	685	OY	553	QDLPRFIVETKRLEVIDIHGARKLESYFEDRWKDKYKGNKNPAOL--	QLEHLRD--	ES	608				
OY	666	SATSIRLOIKKACAGVAGSLs-----	LVISTCKNITSIAWVASPLT--	IEDEBHITSV	735	OY	609	EKTIRLPIFLKLDSTNDFSTMPILTRLLLRNCRIRKL--	POLRPLTNIOIIDACATDI	667	OY	736	TNL-----	KTLSDHLOMORLPGGJLTSLGKMLTKLIMONIMNEDAIKILAEGT	787	OY	668	VEMLEVCLEEKELRILDMKTSIP--	ELADPIADVNLMLLRNCLIE-----	L	718
OY	788	KNLKMCLFHLTHLSDIGEMDYTVKSLSEPCDLEIOLVSCCLSANAVKILAOILHNF	847	OY	719	PSIEK-----	LTHL-----	EVFVSGCIK--	LKNINGSFGEM	748	OY	848	VKLSTLDSBNYLERKDGEALHELIDIMNVLEQLTALMLPWGCDVGOSSLSLKHLEAV	907	OY	749	SYLHEVNLSTENLS-----	ELPDKISTELSNKELTI-----	RKCSKIATLE	789	

Db 708 PENMASLVEFLNKGCTGLESJPKINLRSLKTLILSNCSNL-----EEFVVIS 754
 Qy 735 VNNKLTSLIHQDQNRLEGGLTDSGNLKNLTKLIMDNINKMEEDAIKLAEGKLNKKM- 793
 Db 755 EFLVTLTLDGTALTKLP--QDMVKTLVTKLYMKDCM---LVKLPFEEDKLYKIQ 805
 Qy 794 -----CLFHLTHLSDIGEGMDYI-----VKSLS-EPDCDEEIQVSCCTSA 834
 Db 806 ELVCSGCK-RLLSLPDVKKNNQCLDILLDGTAIRKIPHISLEERLCISRNKISCL--S 862
 Qy 835 NAKVILAO---NLHNLVKL-SIIDLSENYLEKGN-----BALH-- 869
 Db 863 NDRIILSOLKWLDKYCKTKIVSIPELPTNLQCLDANGCESLTTVANPLATHLPFQIHST 922
 Qy 870 -----ELIDRN-----VLEQLTALMLPWGCDVQGSLSLKHLEVPOLYKLG-LKWR 918
 Db 923 FLFTNCDKLDRAKGFEALFSTCFP-GCEV---PSWFCB-EAVGSVTLKLLPLHW- 975
 Qy 919 LDTFIRILG 928
 Db 976 ---NENREVG 982

RESULT 9
 T06403
 resistance complex protein I2C-1 - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 17-Nov-2000
 C:Accession: T06403
 R:Orig. N.; Eshed, Y.; Paran, I.; Presting, G.; Aviv, D.; Tanksley, S.; Zamir, D.; Fluhr,
 Plant Cell 9, 521-532, 1997
 A:Title: The I2C family from the wilt disease resistance locus I2 belongs to the nucleot
 A:Reference number: 215652; MUID:97290204
 A:Accession: T06403
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1220 <ORF>
 A:Cross-references: EMBL:AF004878; NID:g2258314; PIDN:AA863274.1; PID:g2258315
 C:Genetics:
 A:Gene: I2C-1
 A:Map position: 11
 C:Function:
 A:Description: confers resistance against Fusarium oxysporum
 C:Superfamily: disease resistance protein RPS2; leucine-rich alpha-2-glycoprotein repeat

Query Match 3.28; Score 170.5; DB 2; Length 1220;
 Best Local Similarity 19.8%; Pred. No. 0.043;
 Matches 231; Conservative 161; Mismatches 393; Indels 383; Gaps 61;
 Qy 8 SRALQRMGMVYIKQITDDLEFWNVNLANREVNIIICEVEYDQDAAGIIMILKSGSECN 67
 Db 241 TKGILQEGSTDLK--ADD-----NLNQLQVLYKLRADNMQ-----LQVLYKELN 284
 Qy 68 -LFLSKLKE-W--NYPLFDLNGSLFQHTSEGLDD-----LAQDLKDYH 110
 Db 285 GKRFELVLDVDMVNDNYPFMDL--RNLFLQ---GDIGSKIIYTTTKESVALMDSGAIY- 338
 Qy 111 TPSPFNFPYG---EDIDIIENLKSFTFEPVLMRKD-QHHHREQL-----TLNGL- 157
 Db 339 -----MGLISEDSMALEFKRHS-----LEHKDREHPEFEVQKQIADCKGLPL 383
 Qy 158 -LQALQSCITIEGESGKSKSLTLQRIAMLMGSGKALKTKRFVFLRSNAGSLFETL 216
 Db 384 ALKALAGLRKRSSEVDENRILRSET---WELPSC-----SNQILPAL 423
 Qy 217 CDQLIDIGTIRKOTFMAMLLKLRQVFLLDGYNEFKPQNCPEIALIKENHRP-KMV 275
 Db 424 MLSTVDLPAHLKQ-----CFACATLPYPDQPRKQV 455
 Qy 276 IVTTTECLRIHROGA-----LTAEGVDMEQSAQALIREVLLKELAGLLQIQ 327

Db 456 IHLWIANGLVH--QFHSNGQYFIELRSKLEPMASPESEPDVEEFLMHLVNDLQAIASS 513
 Qy 328 SRCRLNLMKTLFLVYITAI-----QMGSE-----FHSITQTL-----EFTFDL 370
 Db 514 NHCIRLEDNKGSHMLEOCRHMSYSIGDGEFEKLSFKESQRLTLPIDIQH-YSKKL 572
 Qy 371 IQKNKHKHGVAAADFISLIDHC-----GDALGVGSHFDELODVSSVNDV- 421
 Db 573 SKRVLHNLPTLRSLRALSTSHYQLEVLPNDFIKLRLRLDLSERSITKLPDSIVLY 632
 Qy 422 -LTTGLL--CYYTQREPKYKFFHKSFOEYTAGRUS--LTSHPPEVYKNGYL- 474
 Db 633 NLETLISSCYLEELPLQMEKLIINRLHIDISNRRKMPHLRLSLQVYGAKEFLVG 692
 Qy 475 -QKWSISDITSTYSSILRYTCGSSV---EATRAVMKHLAAYQHGCLLGISTAKRPLMR 530
 Db 693 GWRMEYGLAEANLYGSLSILELVENVDRREAVKAMRKNHVEQ-----LSE---W- 741
 Qy 531 QESLOSVKNTFEQEL-----KAININSFVECGIHLQESTSKALSQEFEEFQ 581
 Db 742 SESISADNSOTERDLDLDELREKNIKAVEITGY----- 774
 Qy 582 KSLVINSGNIDPYLFD-----FEEHLPNCAS-----ALDEIKL- 614
 Db 775 -----RGTFPNWVADPLFVKLVHLYLRNCKDCYSLPALQGLPCLFELSGMHGIRVY 829
 Qy 615 -DFYGGAMASWEKAEDTGGIHMEAPETTYIPSAVSLFFWKKO-----EFTLEVTL 666
 Db 830 EEFYG--RLSSKKRFPNSLVKRFEDMPG-----WKQWHTLGIGEPFTE- 871
 Qy 667 RDFSRLNKQDITYLGKIPSSATSLRLQRCAGVAGSLVLT-----CKNIYSIM 718
 Db 872 -----KLSIKNCPESLEPIQFSSLSKLRHIDCCKSVTSFP 907
 Qy 719 VEASPLTIE-----DERHITSVNLKTLSDHQLONRPLPGLTDSLGNL 762
 Db 908 FSILPTTLKRIKIGSCPRLKLEAPVGEVFEYLVIDGCVDDISPEFLPARGLSIENC 967
 Qy 763 KNLKRLIMDNINKMEEDAIKLAEGKLNKKCL-----FHLTHSDIGEGMDYIVKSLSS 817
 Db 968 HNVTRFLPTATESLH-----IRNCEKLSMAAGAAQTSLSLNG-----CKRLKC 1013
 Qy 818 EP-----CDLEETIOLVSCCISANAVKILQNLHNLVKLSTIDSENVYKLE--DGNALHELI 872
 Db 1014 LPPELLPSLKEIRLYYC-----PEIEGELPN---LQILDI--RYCKKLVNGKREMH--- 1059
 Qy 873 DRNVLEQLTALMLPW-GCDVQGSLSLKHLE--EVPQVLYGLKKNRLLDTFIRILGA 929
 Db 1060 -----LQRLTELMIKHDGSD-----EHIEHMELEPSIO-----RLFTFNLKTLSS 1099
 Qy 930 FFGKNPLKNPQOLNLAGN--RVSSDGMALAFMGVEENLKOLVFPFSTKEFLPPALVRL 987
 Db 1100 QHLKS-LTSLQFLRTIVGNLSQFQSGQLSSFSHLTSLQTLQIMFNLQSLPESALPSSL 1158
 Qy 988 SQ-VLSKTLFLEEARLVGWGCFDDDDLSV 1014
 Db 1159 SHLIISNCPNLQSLPLKGMPSLSLSTLSI 1186

RESULT 10
 T15864
 hypothetical protein C56E6.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 R:Fullon, L.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C56E6.
 A:Reference number: S69019
 A:Accession: T15864
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1066 <FUL>

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, Nature 402: 761-768. 1999

[illegible]

```

Db      719  NLRSGELILSGCSKLFPE-----LKEFWKSIKILL-----DGTAIKOMPILL 764

```

Db 765 OCIOGSHSVANKTLPNSLSDYLLPSSLLSLCLSGNDIESLHANSIQYLKMLDKNCK 824
OY 858 -----NYLEKGNENALHEL----- 871
Db 825 KIKSVSLVPLNKLCKDAHCCDLSLEVGSPFLAVMTGKIHCTYIFTNCKIKLDQVATESNI 884
OY 872 -----IDRMN---VLEQTLTALMLPWGCDVQGS 895
Db 885 SFTWRKSQMSDALNRKNGFVLESVSTCFP-GCEVPAS 923

RESULT 12
T10504
disease resistance protein Cf-2.1 - currant tomato
C:Species: Lycopersicon pimpinellifolium (currant tomato)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T10504; T10515
R: Dixon, M.S.; Jones, D.A.; Keddie, J.S.; Thomas, C.M.; Harrison, K.; Jones, J.D.
File: the tomato Cf-2 disease resistance locus comprises two functional genes encoding
Reference number: Z17062; MUID:96190812
A:Accession: T10504
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1112 <DIX>
A:Cross-references: EMBL:U42444; NID:91184074; PIDN:AC15779.1; PID:91184075
A:Experimental source: cultivar Cf 2
A:Accession: T10515
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1066, '1', 1068-1085, 'E', 1087-1110, 'R', 1112 <DIX>
A:Cross-references: EMBL:U42445; NID:91184076; PIDN:AC15780.1; PID:91184077
A:Experimental source: cultivar Cf 2

Query Match 3.1%; Score 165; DB 2; Length 1112;
Best Local Similarity 21.6%; Pred. No. 0.08;
Matches 139; Conservative 81; Mismatches 199; Indels 224; Gaps 31;

OY 425 GLLCKYTRQRPKPKYFKHKSFOEYTAGRSLSLTSHPEEYTKNGYLOKMWISDIT 484
Db 140 GLAALQIIR-----IFHNQ-----LNGFIPEKI-----GYLSLTKLS-LG 175
OY 485 STYSSLAYTGGSSVEATRAVWKHLAAVYQHCGCLLSIAKRPMLROSLQSVKTTQOE 544
Db 176 INFUS-----GSLPASVGNLNNLSFLYLYNNQLSGI-----PEISYLRSLTELD 221
OY 545 ILKAININSFVECGIHLQESTSKALSQEEFA-----FFQKSLYINSGNIPD 593
Db 222 L-----SDNALNGSIPASLGNNNNLSFLYLYGNQL-----SGSIPE 257
OY 554 ---YV-----FDPEHLPN---CASALDFIKLDF---YGAWASWEKAEDTGGIHMEEA 639
Db 258 EICYRLSLTYLIDLSENALNGSIPASLGNNLSFLYLYGNQLSG---SIPEIG----- 308
OY 640 PETYIPSAVSLFFEMWKEFTLEVTLRDFSKLNQDITTYLCKITFSSTSLRLQIKRCAG 699
Db 309 ---YV-----RSLANVLGSENALN----- 324
OY 700 VAGSLSLVSTCKNITSYLMVEASPLTIEDERHITSVTNKLKLSIHLDONORLPGLTDSL 759
Db 325 ---GSLPASLGNNKNSRLNLYNNQLSGSI---ASLGNNLNSMLYLYNNQLSGSI PASL 379
OY 760 GNLKNLYTLINDNIKANEEDAIKLAESLKNLKKMCLFHLTHLSDIGEGMDYIVK---SL 815
Db 380 GNLNNLSMLYLYNNQLSGSI PASLG---NNLNNLSRLYLYNNQLSGSIPEIGYLSLTLYDL 438
OY 816 SSEPCD-----LEEIOLVGCCLSANVKKLLAQVLAHLVLTSLDISEN 858
Db 439 SNNSTNGITPASFGNMSLAPFLYLENOLASS-----VPEIGYLRSLNLYNDLSEN 489
OY 859 YLEKDGNEALHELIDRMNVLEQTLTALMLPWGCDVQGSLSLLKHLSEVPQVTKLGKLMWR 918

Db 490 AL-----NCSIPASFGNLLNNLSRLN-----VNNQLSGSIPEEELGYLRSLNLYNDL-- 533
OY 919 LTDEIR-ILGAFGCKNPLKNFQQLNLAGNRVSSD-----GWL----- 955
Db 534 LSENALNGSIPASFGN---LNNLSRLNLYNNQLSGSIPEIGYLRSLNDLGLSENALNGSI 591
OY 956 -AFMGVFENLQVLFDEFTKEFLPDPALVRKLSQVSKLFL 997
Db 592 PASLGNNLNSMLYLYNNQLSGSIPEI-----GYLSLTYL 628

RESULT 13
T06145
disease resistance protein homolog F24J7.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06145
R: Bevan, M.; Vitale, D.; Liguori, R.; Argyriou, A.; De Simone, V.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15493
A:Accession: T06145
A:Molecule type: DNA
A:Residues: 1-1715 <BEV>
A:Cross-references: EMBL:AL021768; GSPDB:GNO0062; ATSP:F24J7.80
A:Experimental source: cultivar Columbia; BAC clone F24J7
A:Gene: ATSP:F24J7.80
A:Map position: 4
A:Introns: 4/1; 122/2; 476/3; 575/3; 1088/3; 1154/2; 1512/2; 1610/1; 1662/1

Query Match 3.1%; Score 164.5; DB 2; Length 1715;
Best Local Similarity 20.3%; Pred. No. 0.16;
Matches 217; Conservative 134; Mismatches 312; Indels 407; Gaps 56;

OY 53 GIIMILKSGSECNLFPSL-----KENYPLFO---DLNGOSL-PPHOSGRLD-DLA 102
Db 76 GLTNSIVK---HC-LELKNYPPDKVDENRNALMDLADLGGVSHKRSDELVEKIV 131
OY 103 QDLKDLVHTPSFLNFPLEGEDIDIFNLKSTFTEPVLMRKDQHHRYEQLTLNGLQALQ 162
Db 132 ADVROKIDRRRIGVY-----SRLTKIETYLCKQ 160
OY 163 SPCTIE-----GEGKQKSTLLQRIAMMGSKKALTKFVFYFLKLSR----- 207
Db 161 PGCIIRSLGIMGAGIKGTTLAR-----AAVDOLSRDFEASCFIED 201
OY 208 -----AAGLPEPTCODOLDIPGTRKQTFMAMLLK-LR-GRVFLLDGVNEFRPQNC- 259
Db 202 FDRPEQKFGFLLEKQL-----GVNPOVTRLSILKTLRSKRILLVLDVNR--KPLGATS 255
OY 260 ---EIEALIKENHKKNNVIVTTTECLRHIRQEGALTAEGMDTEPSAQLIREVLIKE 316
Db 256 FLCEFDMLGP-----GSLIIVTSQDKQVLYCQVNEIKYVGLKMHSLQLFSCARGKD 310
OY 317 LABGLLQIQK-----SRCLRNLM-KTPL---FVYTCAIOMGESFPHSTQ 359
Db 311 VPQNLLELSMKFPVDYANGPNLALSLCKNKGKTPIDMKSVL-----ELKRLHS 361
OY 360 TTLF---HTFDDLILQKNKH-----KHKGVAASDFIRSLDHG---DLALEGVFSGHKF 406
Db 362 DKITVYKLSSTDALSVSEKELFLDIVFFFRGANDVWQSLAGCGFPRPRGIEALVDKSF 421
OY 407 DPELDVSSVNEVDVLTTLG-----LCKYTAQREPKYKFFHKSFOEYTAGR-- 453
Db 422 VTVAENRQVNA-NLIVDGLKIINDQSEIDCMCYRFVDANSQSLIKHKEIRESEQYED 480
OY 454 -RLSSLSLSHPPEEYTKNGYLOKMWISDITTYSS----- 489
Db 481 VKAINLDSNLP---FQGHIAFOHMYNLYLT--TYSSINPTKPDLPGLPGDPELPPELR 536
OY 490 LLRTCGSSVEATRAVWKHLAAVYQHCGCLGLSLAK--RPLWRQESLQSVKNTTQETLIKA 548

```

Db 537 LHMTC---YPLHSPQNG--FOYLVELNMCSSKLKLM-----GGTNLEVLKR 582
OY 549 ININSVEGCIHLHYOESTSKSALSQEFAPFOGKSLYINGNIPDYLDFEFLPNCASA 608
Db 583 ITLS---CSVOLLN-----VDELOQSPNIE-- 604
OY 609 LDEIKLDFYGAMASWEKAEDTGGIMHEPAPETYIPRAVSLFFNWKQOFRTLEVTLRD 668
Db 605 ---KIDL-----KGCLELOSPDPT----- 620
OY 669 FSKLNKODITYLKITSSATSLRLQIKRCGAVAGSLV--LSTCKNTYSILMEASPLTIE 727
Db 621 -----GOLQHLRIYDLSTCKTKKSF--PKVPISIR 648
OY 728 DERHI---TSVTNKLTLSDHDLQNRPLGGLTDSLGNLKNLTKLIMDKNNEDEAIKLA 785
Db 649 -KHLIGCTGIRDSLIN-HSESOQLTKRL-----ENVSSNDHKRQVL 691
OY 786 GLKNLKMCLFHLTHLSDIGEGMDYIVKSLS---SEPCOLEBIQ-----LVSCCLISANA 836
Db 692 KLDSS---HLGSLPDI-----VIFESLEVLDGSGSELEDIOGFPQMLKRLYLAKTA 741
OY 837 VKTLAQL--HNLVKLSTLDSENVLEKDGNEALHELLDRNNVLEQLTALMLPGCDVQGS 895
Db 742 IKVEPSSLCHHISKLVKLDN-----ENCERLRLPMGSKNMKYLAVALKLS--GC----- 788
OY 896 LSSLKHLKEVPOLVKLGLKNMRLTDEIRILGAFEGKNLKNFOOLNLAGNVSDDGWL 955
Db 789 --SNLENIKRLP-----RNLKEIYLXAGTAVKE----- 813
OY 956 AFNG--VFENIKQLVFEDFSTKEFLPDPAVLKLSQVLSKLTPLOEARLVG 1004
Db 814 -FPSTLETLSEVLLD-----LENCKKIQGLPTGMSKLEFLVMLKLSG 856

```

RESULT 14

T30553
 disease resistance protein Hcr2-5D - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
 C:Accession: T30553
 R: Dixon, M.S.; Hatzixanthis, K.; Jones, D.A.; Harrison, K.; Jones, J.D.G.
 Plant Cell 10, 1915-1926, 1998
 A:Title: The tomato Cf-5 disease resistance gene and six homologues show pronounced allelic variation
 A:Reference number: 220855; MUID:99030197
 A:Accession: T30553
 Status: preliminary; translated from GB/EMBL/DDBJ
 Molecule type: DNA
 Residues: 1-1016 <DI>
 A:Cross-references: EMBL:AF053998; NID:93894392; PID:93894393; PIDN:AAC78596.1
 C:Genetics:
 A:Note: Hcr2-5D

```

Query Match          3.1%; Score 164; DB 2; Length 1016;
Best local Similarity 22.1%; Pred. No. 0.08;
Matches 147; Conservative 115; Mismatches 222; Indels 182; Gaps 38;

OY 459 LTHHEPEEYKKGGLQKMWIS-DITSYSSILRYTCSSVETRAVMHIAAVYOHG 517
Db 203 LSGFTPEET---GYLKSILTKLSDLN-----FLSGSIPASLGNLNNLSFLYLYNN 249
OY 518 LIGLSIAK---RPLWR-----QESLOSVKNTTEOEI----- 545
Db 250 QLSGSIPEELGYLSRLTKLSLGINFLSGSIPASLGNLNNLSRLDLYNNKLSGISPEIGY 309
OY 546 LKAININSEVEGCIHLHYOES--TSKSALSOEFAPFOGKSLYINGNIPD--YL----- 595
Db 310 LRSLL---TYLDLGENMALNGSIPASLGNLNNLSFLYLYNNQL---SGSIPPEIGYLSRLTY 363
OY 596 FDFEFLHPN-----CASALDFIKLDFYGAMASWEKAADDTGGIMH-----EAPRT 642

```

```

Db 364 LDIGENALNGSIPASLGNLNNLSRLDLYNNKLSG--SIPPEIGYLSRLTYLDIGENALNG 421
OY 643 YIPRAVSLFFNNKQOFRTLEVLRDFPSKLNKODITYLKITSSATSLRLQIKRCAGAV 702
Db 422 SIPASLGNL-----NNLFMLYLYNNQLSGISPEIGYL-----SSLTEIYIGNN---SLNG 469
OY 703 SLSVLSTCKNIYSILMEASPLT---IEDERHITSVTNKLTLSDHDLQNRPLGGLTDSL 759
Db 470 SIPASLGNLNNLSFLYLYNNQLSGISPEIGYLSLTEL-----FLGNNSLNGSIPASI 523
OY 760 GNLKNTLTKLIMDKNNEDEAIKLAEG-LKNLKMCLFHLTHLSD--IGEGMDYIVKSL 815
Db 524 GNLLNLSRLYLYNNQLS--GSIPASFGNNKRLDTL-----FLSDDDLIGELIPSCNLT 575
OY 816 SSE-----PC--DLEIQLVSCCLISANA--V--TLAQLNHLNVLKSLTLDLSEN 858
Db 576 SLEVLVMSRNNLKGVPQCGNISDLHLS--MSSNSPREFLPSSNLSLTKLIDFGNR 633
OY 859 YLEKD-----GNEALHELLDRANVLEQLTALMLPGCDVQGSLSLKH---LEEVP-- 907
Db 634 NLEGAIPQEFNGNISLQVPMQN--NKLSG-TLPTNFSIGCSLISNHLGNELADEIPRS 690
OY 908 -----OLVXKGLKNMRLTDT-----EIRLGAFFGK--NPLK-----NEOOL 942
Db 691 LDNCKKQLVLDLSDNQNDLTFPWLGLTPELRVRLTSNKHGPIRSSGAEIMFPDLRII 750
OY 943 NLAGNRYSSDGMVLAFGVFNELKQLVFFPSTKE-----FLPDPAV-----RKLQVYL 991
Db 751 DLSRNAFSQD--LPLSLFELHAKMRTVKTMEPEYESYDSDSVVYVKGLELEIVRLI 807
OY 992 SKLTFPL 997
Db 808 SLVYII 813

```

RESULT 15

A40264
 kinesin-related protein Eg5 - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
 C:Accession: A40264; S18764
 R:Le Guellec, R.; Paris, J.; Couturier, A.; Roghi, C.; Philippe, M.
 Mol. Cell. Biol. 11, 3395-3398, 1991
 A:Title: Cloning by differential screening of a Xenopus cDNA that encodes a kinesin-related protein
 A:Reference number: A40264; MUID:91246212
 A:Accession: A40264
 A:Molecule type: mRNA
 A:Residues: 1-1060 <LE>
 A:Cross-references: EMBL:X54002; NID:964869; PIDN:CAA37950.1; PID:964870
 A:Note: The authors translated the codon GCT for residue 784 as Leu
 C:Superfamily: kinesin-related protein Eg5; kinesin motor domain homology
 C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
 F:12-358/Domain: kinesin motor domain homology <KMO7>
 F:98-105/Region: nucleotide-binding motif A (P-loop)
 F:104/Binding site: ATP (Lys) #status predicted

```

Query Match          3.1%; Score 164; DB 1; Length 1060;
Best local Similarity 20.3%; Pred. No. 0.085;
Matches 214; Conservative 155; Mismatches 383; Indels 302; Gaps 52;

OY 24 TDLLFVNVNINREVINICEKVEDAARGIT---HMLIKGSESCNLF--LKSLEK 76
Db 114 SDEEFTW-----EDPRLAGIIPRTLQHFKEKLSNGTFSVVSLEIY 157
OY 77 NYPLFQDLN-----GQSL-----FHQTSFG--DDDLADQLKD-LYN-----T 111
Db 158 NEELFLLSLSPDVGRLQFDDPRNKRGVYIKGLEISVHNKDEYVNLIERGAARRKA 217
OY 112 PSFLNPLRGED--DILFNLKSTFERYVLMKRDQHNHREQLTLNGLQALQSPCTIEG 169
Db 218 STLMAVSSSHSVSYTTHMKETVVG-----EELVKGKGLNLYD-----LAG 261

```


THIS PAGE BLANK (USPTO)

[illegible]

453	rgArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys	466
2656	TCAGCTCATTTGAACTCTCGATTGATAGACGACGACATCAAGATTGG	2705
470	GlyAsnGlyTyrLeuGlnLysMetValSerLieserAspLietrSerTh	486
2706	GGACGTGATCATTTTAAACAAATCAACACACCCATGATGATGTAAGGC	2755
486	tyrSerSerLeuLeuArgTyrThrCysGlySerValGluAlaIra	503
2756	CTACACAACTTTTGTGAACATACTGTC.....TTCAGCTCTCTTCACAA	2799
503	rgAla.....ValMetLysHisLeuAlaAlaValTyrGlnHisGly	516
2800	AAGCAGGCGCCCAAAATTGTGTCTCATTTGCTCCATTTAGTGGATAC...	2846
517	CysLeuLeuGlyLeuSerLleAlaLysArgProLeuThrArgGlnGlu	533
2847AAAGAGTC	2854
533	rIleuGlnSerValLysAsn.....	539
2855	ATTGGAGATATATGTGAAATGATGACTACTTAAAGCAGCCAGACA	2904
540	...ThrThrGlnGlnLleLleLeuLysAlaIle.....AsnIle	551
2905	TTTCACTGCAGATGCGATTACTACTTAGGGGATTGTGGCAAAATTGTGCACA	2954
552	AsnSerPheValGluCysGlyLleHisLeu.....	561
2955	GCTTACTTTCCAATGGTTTGCACAAACATTTACTGTTCTTGCCCTGAAAC	3004
562	...TyrGlnGlnSerThrSerLysSerAlaLeuSerGlnLupheGlu	577
3005	TGCTTATCAACACACT...GTTCGCGCGTTCCTTCATTTGTTGGC	3051
577	lApheGlnGlnGlyLysSerLeuTyrLleAsnSerGlyAsnIleProAsp	593
3052	AATTCCTTCACAGGAGAACACTGACTTTGGTGCGCTTAACCTTA...CAG	3098
594	TyrLeuPheAsp.....	597
3099	TACTTTTTCGACCCACCCAGAAAGCTTGTCATTTGGAGAGACATCCACT	3148
598PheGlnHisLeuP...603	
3149	CTCAATACGAGGAATAAGACATCACCCAGACACACTTTTTCAGTTGCT	3198
603	roAsnCys.....AlaSerAlaLeuAspPheLleLysLeuAspPheTyr	617
3199	AAACATGTTTGCACAAATCACAGGCGCAACTATATGATCAGGACTATGCT	3248
618	GlyGly.....AlaMetAlaSerTrpGlu.....	625
3249	TCTGCCTTTGAACCATATGATGAATGAATGGAGGGAATTTAGCTGAAAAAGA	3298
626LysAlaIleGluAspT...631	
3299	GGATATATGTAAGAGCTATATGATATGACAGCGAGGCGATCACACAGCA	3348
631	hrcGlyGlyLleHisMetGluGluAlaProGluThrTyr...IleProSer	646
3349	TTAGTACTGCTGCTATTGGAAACTTCTCCAAAGACACTACAAGATTCC...	3395
647	ArgAlaValSerLeuPheAsnTrpLysGlnLupheArgThrLeuGlu	663
3396TCGCTAGA	3403
663	valAlaThrLeuAlaGAspPheSerLysLeuAsnLysIleAspLietrTyrL	680
4404	ACTGCATGTGAATGATATTGATGTTGTAGGCGCAGATATGCTTGAGATTC	3453

[illegible][illegible]

136 GluPro..... 137
 749 TCACCAACAAAGAACTAGACACCTTCAACTCTTGGTTGATGATTC 798
 138 ...ValLeuTPArgLysAspGlnHisHisArgValGluGlnLeu 153
 799 TGGTATCTTTGGAAGAAATGAAATAGAAATTTGGTGGCGTTGT 848
 153 hrLeuAsnGlyLeuGlnAlaLeuGlnSerProCysIleIleGluGly 169
 849 TGCTATGATGATACAAAGCAAAAAATGGCTGATGTTCTATTGTGGGA 898
 170 GluSerGlyLysGlyLysSerThr..LeuLeuGlnArgIleAlaMetLeu 186
 899 ATGGCGCGCATGGGTAGACACACTGTTAAAGCCGTTTACATGATG. 947
 186 rpeLysGlyLysCysLysAlaLeu.....ThrLysPheLys 198
 948AGAGAGTCGAGAAACATTTTGGTTGCACAGCTTGTTGTG 989
 199 PheValPhePheLeuArgLeuSerArgAlaGlnGlyLeuPheGluTh 215
 990 TTTCTGAGCATATGATGCTTTTCAGAAATACCAAAAGCTTACTTCAAGAA 1039
 215 rLeuCysAspGln.....LeuLeuAspIleP 224
 1040 ATTG...GATCACTGACTGGAAGCTGATGACAACTTATCATGCTACA 1086
 224 roGlyThrIleArgLysGlnThrPheMet..AlaMetLeuLysLeu 240
 1087 AGTCAAAATTAAGGCTGATGACAACTTAAATCAAGTCAAGTCAAAATGA 1136
 240 rGln.....ArgValLeuPheLeuLeuAspGly...Tyr 250
 1137 AGCAAAAGCTGATGAAAAAGTTTCTTGTGCTTGAATGACGTGTGG 1186
 251 AsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuLysGly 267
 1187 AATGAT.....AATTAATCTGAGTGGGATGACTTG..... 1216
 267 uAsnHisArgPheLysAsnMet.....Y 275
 1217AGAAATCTTTTTCACAGGGGATATAGAGATAGA 1253
 275 allLeuValThrThrThrGluCysLeuArgHisIleArgGlnPheGly 291
 1254 TCATTGTAAAGACACGTAAAGAGGTGTGCTTGATGATGATGATGGG 1303
 292 AlaLeu..... 293
 1304 GCAATCTACATGGGAATTTCTGTAGTGAAGACCTTGGGCTATTCAA 1353
 294ThrIleG 296
 1354 ACGACATTCATTAGAGCACAGATCCCAAGAAATCCAGAAATTTGAAG 1403
 296 LuValGlyAspMetThrGluAspSerAlaGlnAlaLeuIleArgGlyVal 312
 1404 AGGTGGAAAAACAATTCACAGACAGTGCAGAGGTTG.....CCTTTA 1447
 313 LeuIleLysGluLeuAlaGlnGlyLeuLeuGlnIleGlnLysSerAr 329
 1448 GCTCTAAACACTTGTCT...GGTATGTACGCAAAATCAGAGGTGGA 1494
 329 gCysLeuArgAsnLeuMetLysThrProLeuPheValIleLeuThys. 345
 1495 TGAGTGGAGAAACATTTTACGAGGTGAATGTGGACCTTCCAGTTGTT 1544
 346AlaIleGlnMetGlyGlnSerGluPheHisSer 356
 1545 CGAATGATATATACACGCGCTATGTGAAGCTTCAATGATCTCCCTGCA 1594

357 HisThrGlnThrThrLeu..... 362
 1595 CATTTAAAGCAATGTGGCTTATTGTCAATATATCCAAAGATATCA 1644
 362 362
 1645 ATTTGCAAAAGACAGACTTATTCACCTGTGGATTCGTAATGCTGTAC 1694
 363PheHis.....Thr 365
 1695 ATCAGTTTCATTCGGGTAAACCAATACCTTTATCGAGTTGAGATCAAGATCA 1744
 366 PheTyrAspLeuLeuIleGlnLysAsnLysHisLysHisGlyValAl 382
 1745 TTGTTCAAAATGGCCCTCAGACCTTCTGAAGAGACGTAGAGAAATCTT 1794
 382 aAlaSerAspPheIleArgSerLeu.....AspHisC 393
 1795 AATGCATGACCTTGTCAATGATTTGGCAAAATTCATCTTCAATCATTT 1844
 393 ys.....GlyAspLeuAlaLeuGluGlyVal... 401
 1845 GTATAAGCTTGAAGATATACAAAGATCGCATATGTGGAACATGTGCG 1894
 402PheSerHisLysPheAspPheGluLeuGlnAspValSer 415
 1895 CACATGCTCTTATTCATAGACAGACAGATGCTGAGTTGAGAAATGAATC 1944
 415 rValAsnGluAsp.....ValLeuLeuThrThrGlyLeuLeu 428
 1945 ACTTTTAAATCAGACAGCTGAGGACATTAATCTTCCATGATATCCAGT 1994
 428 yLysTyrThrAlaGlnArgPheLysProLysTyrLysPhePheHisLys 444
 1995 TCCATTCTCAAAAAAACTAAGCAAG.....AGGCTGTGCATTAAC 2035
 445 SerPheGlnGluTyrThrAlaGlyArgArgLeuSerLeuLeuThr 461
 2036 ATACTGCTACACTA.....AGATCTGTAGGCGCACTATCATTTGTC 2076
 461 rHisGluProGluGluValThrLysGlyAsnGlyTyrLeuGln...LysM 477
 2077 TCATTACCAATTTGAGGTGTTGCCAAATGACTTGTATATCAATTAAGC 2126
 477 etValSerIleSerAspIleThrSerThr..... 486
 2127 TCCTCAGATTTTGGACCTTCTGACACATGATTAACAAGTCCGGAT 2176
 487TyrSerSerLeuLeuArgTyrThrCysGly 496
 2177 TCCATTTTGTGTGTATTAATCTTGAAGACACTTCTCCTGCATCTGTGA 2226
 496 ySerSerValGluValThrArgAlaValMetLys..... 507
 2227 ATATCTTGAGAGCTACCGCTGCAGATGAGAGAGTTGATTAATCTGGCTC 2276
 508HisLeuAla 510
 2277 ATCTTGACATAGACACACACGCGGCTTGAAGATCCCACTACATCTAGC 2326
 511 AlaValIleGlnHisGlyCysLeuLeuGlyLeuSerIleAlaLysArgPr 527
 2327 AGGTGAAGCCCTCCAAAGTGTGTGGTGAGCCCAAGTCTTGTAGAGTGG 2376
 527 oLeuTyrArgGlnGluSerLeuGlnSerValLysAsnThr...ThrIleG 543
 2377 T...TGGAGATGAGATATTTGGGTGAAGCACCAACTTATATGATCTC 2423
 543 InGluIleLeuLysAlaIleAsnIleAsnSerPheValGluCysGlyIle 559
 2424 TATCAATTTATGACTTGGAAAAATGTGGTATAGAGGAAGCTGTGAAG 2473
 560 HisLeuTyrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheGly 576

```

2474 GCACAGATGAGGAGAGATCATCTTGAGCAATATCATTTGAGTGAG 2523
576 uAlaPhepHeGIngly..... 581
2524 TGAACCATTTAGTGTGACAAATTCACAAACAGAACATACCTTGATG 2573
582 .....LysSerLeuTyTle.....Asn 587
2574 AGCTACGCCACATAAAACATTAAAGCAGTTGAATCAGTATATAGA 2623
588 SerGlyAsnIleProAspTyTleuPheAsp..... 597
2624 GGGACAAACTTTCACAACTGGAGTACCTCTTGTGTAAAGCTGTG 2673
598 .....PhepHeGInHisLeuProAsnGlyAsn..... 607
2674 GCATTTGTATCTTAGAACCTGCAGAGACTGTATCTCTTGCCAGCAG 2723
608 .....AlaLeuAspPheIleLysLeu..... 614
2724 GACAACTCCCTGTTTGAATTCCTTCATTAGAGGATGCAATGGGATA 2773
615 .....AspPheTyTleGlyAlaMetAlaSerTrpGluLys 626
2774 AGAGTGTGACAGAGAGTCTATGAC.....AGATGTCTCTCCAAAA 2817
626 sAlaAlaGluAspThrGlyGlyIleHisMetGluGluAlaProGluThr 643
2818 GCCTTTAACTCTGTGAAAGCTAGATTTGAAGATATGCTCTGAA.... 2863
643 yIleProSerArgAlaValSerLeuPheAsnTrpLysGln..... 657
2864 .....TGGAAAGCAATGGCAC 2878
658 .....GluPheArgThrLeuGluValThrLeuArgAsp 669
2879 ACATAGTAATTTGAGAGTTCCTACACTTGAG..... 2911
669 eSerLysLeuAsnLysGlnAspIleThrTyTleGlyLysIlePheSer 686
2911 ..... 2911
686 eAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyValAlaGly 702
2912 .....AACTTTCATTTAAATTTGCCCTGAGCTCAGCTTG 2947
703 SerLeuSerLeuValLeuSerThr.....Cys 711
2948 GAGATACCATTCCAATTTCAAGTTTAAAGTTAGATATATGTCATG 2997
711 sLysAsnIleTyTleSerLeuMetValGluAlaSerProLeuThrIleGlu 727
2998 TAAGTCTGTACCTCTCTCTTTAGCACTACCTACCTGCAAA 3047
728 .....Asp 728
3048 GATAAAGATATCTGTTGCCCAAAATTCGAATTTGAGGCGCAGTGTGT 3097
729 GluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis 745
3098 GAGATGTTTGTGAGTATTTGAGTGTGATTTGATGTTGTTGTAGATGA 3147
745 pleuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuGlyAsn 762
3148 TATATCACCTGTGAGTTTCTCCACACAGCAGTCAATTTGATATTTGA 3197
762 eLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsnGlu 778
3198 GCCAACAGCTTACAGTTTGTGATTCCTACCTGCAACAGTCTCCAT 3247
779 AspAlaIleLysLeuAlaGluGlyLeuLysAsnLeuLysLysMetCys 795

```

```

3248 .....ATTCGGAATTTGAAAAACTCTGAT 3273
795 u.....PheHisLeuThrHisLeuSerAspIleGlyGlu 807
3274 GGCATGTGAGAGCGCGCCAGCTGACGTCATGAAATATTTGGGA.... 3319
807 LysMetAspTyTleValLysSerLeuSerSerGluPro.....Cys 820
3320 .....TGTAAAGAACTCAAGTCTTCACAGACCTCTTCCA 3355
821 AspLeuGluGluIleGlnLeuValSerCysLysSerAlaAsnAlaVal 837
3356 TCTCTCAAGAACTGCGAGTACTTATGT.....CC 3387
837 LysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIleLeu 854
3388 AGAAATAGAAAGGAGAAATGCCCTTCAT.....TTACAAATCTCG 3428
854 sPheSerGluAsnTyTleGluLys.....AspGlyAsnGluAlaLeu 868
3429 ATATC.....AGATATTCGCAAGAACTGTGATGCGCCGAAAGAGTGG 3472
869 HisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAlaLeu 885
3473 CAT.....TTACAGAGACTCACAGAGTTATG 3498
885 LeuProTrp...GlyCysAspValGlnIleSerLeuSerLeuLeu 901
3499 GATCAAAACATGATGGAGTGAC.....G 3521
901 yHisLeuGlu.....GluValProGlnLeuValLysLeuGlyLys 915
3522 AACATTTGACATTTGGAGTTCCTCTCTATTCAG..... 3559
916 AsnTrpArgLeuThrAspThrGluIleArgIleLeuGlyAlaPheMet 932
3560 .....AGACTATTCATATTCATCTGAACATCTAAGCAGCAACATCT 3603
932 yLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsn... 947
3604 CAAGAC...CTACCTCTCTCAATTTCTACGATATTTGTAATTTAT 3650
948 ..ArgValSerSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsn 963
3651 CTCAGTTTACGTCACAGGCCAACTTCTCTCTTCTCACCCTCAGTTG 3700
964 LeuLysGlnLeuValPhePheAspPheSerThrLysGluPheLeuProAs 980
3701 CTTCAAACTCTCAAAATCTGAATTTCTTAATCTTCATACCTACCTGA 3750
980 pProAlaLeuValArgLysLeuSerGln...ValLeuSerLysLeuThr 996
3751 ATCAGCACTGCCCTCCCTCTCTCAGCTGATCACTCAATTTGCCCTA 3800
996 heLeuGlnGluAlaArgLeuValGlyTrpGlnPheAspAspAspLeu 1012
3801 ATCTCCAACTCTTCATTTAAAGGATGCTCTTCCCTCTCTACACTA 3850
1013 SerVal 1014
3851 TCAAT 3856

```

```

seq_name: /cgn2_6/ptdata/2/ina/6B_COMB.seq.us-09-180-439-1
seq_documentation_block:
; Sequence 1, Application US/09180439
; Patent No. 622532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatixanthis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

```

```
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180,439
CURRENT FILING DATE: 1998-12-06
EARLIER APPLICATION NUMBER: PCT/GB97/01249
EARLIER FILING DATE: 1997-05-08
EARLIER APPLICATION NUMBER: GB 9609681.3
EARLIER FILING DATE: 1996-05-09
EARLIER APPLICATION NUMBER: GB 9619924.5
EARLIER FILING DATE: 1996-09-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 3979
TYPE: DNA
ORGANISM: Lycopersicon esculentum
US-09-180-439-1

alignment_scores:
  quality: 172.00      length: 639
  ratio: 0.533         gaps: 38
  percent similarity: 50.548      percent identity: 23.005

alignment block:
US-09-697-089-2 x US-09-180-439-1 ..

Align seg 1/1 to: US-09-180-439-1 from: 1 to: 3979

459 LeuhrSerHisGluProGluGluValThrLysGlyAsnGlyTyrLeuG1 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1260 CTTTCTGCGCTTATTCCTGAGAAATA.....GGTTACCTTAG 1297
475 nlyMetValSerLeuSer...AspIleThrSerThrTyrSerSerLeuL 491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1298 GTCCTTACCTAGCTACTCTTGATATCAAC..... 1328
491 euatGlyThrCysGlySerSerValGluAlaThrArgAlaValMetLys 507
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1329 .....TTTCTTAGTGCTTCCTCTGCTCATTTGGGGAATCTGAC 1370
508 HisLeuAlaAlaValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAl 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1371 AACTTGCTTTTGTATCTTTACATATAGCTTTCTGCTCTATT. 1418
524 alyArgProLeuThrParGlnGlnSerLeuGlnSerValLysAsnThrT 541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1419 .....CCTGAAGAATAGTTACCTAGAGTCACATTA 1449
541 hrGluGlnGluIleLeuLysAlaIleAsnIleAsnSerPheValGluCys 557
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 CTAG.....CTATCTTTG 1463
558 GlyIleHisLeuTyrGlnGluSerThrSerLysSerAla..... 570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1464 GGTATCACTTTCTTAGTGCTTCCTGCTTCATTTGGGGAATCTTAA 1513
571 .....LeuSerGlnLupheGluAlaPheheGlnGlyLysSerLeuTyrI 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1514 CAACCTGTCTAG.....TTGATCTTTTCA 1539
586 leAsn.....SerGlyAsnIleProAsp.....TyrLeu..... 595
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1540 ATAATAGCTTTCTGCTCTATTCCTGAAGAATAGTTACCTAGAGCTCT 1589
596 .....PheAspPheheGluHis.....LeuProAs 604
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1590 CTTACTTACCTAGATTGGGAGAAATGCTTTATGCTTATTCCTTC 1639
604 nCysAlaSerAlaLeuAspPheIle...LysLeuAspPheTyrGlyLys 620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1640 TTCTTGGGGAATCTAAACACTTGTCTAGTGTGATCTTACATATA 1689
620 IametaIaSerTTrpGluLysAlaIaGluAspThrGlyIleHisMet 636
```

```
1690 AGCTTTCTGC.....TCATTCCTGAAGAATAGTTACCTAGAGCT 1733
637 .....GluGluAlaProGluThrTyrIleProse 646
1734 CTTACTTACCTAGATTGGGAGAAATGCTTTATGCTTATTCCTGC 1783
646 rArgAlaValSerLeuPhePheAsnTrpLysGlnLupheArgThrLeuG 663
1784 TTCAATGGGGAATCTG.....ACAACCTGTTATGTTGT 1818
663 luValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyr 679
1819 ATCTTTACAAATATACGCTTTCGCTCTATTCCTGAAGAATAGTTAC 1868
680 LeuGlyLysIlePheSerSerAlaThrSerLeuArgGlnIleLysAr 696
1869 CTG.....AGTTCTTACTGAACTATATTTGGGTAATAC. 1904
696 gCysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysA 713
1905 .....TCCTTATAGGCTCTATTCCTGCTTCAATGCGGAATCTGAACA 1947
713 snIleTyrSerLeuMetValGluAlaSerProLeuThr.....Ile 726
1948 ACTGTTTATGTTGATCTTACATATATACGCTTCTGCTCTATTCCT 1997
727 GluAspGluArgHisIleThrSerValThrAsnLysThrLeuSerIle 743
1998 GAAGAATAGTTTACCTGAGTCTCTTACTGAACTA..... 2033
743 eHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuG 760
2034 .....TTTGGGTAATACCTCTTAATGCGCTATTCCTGCTCTATTCCT 2079
760 lyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsn 776
2080 GGAATCTAAACACTTGTCTAGGTTGATCTTACAAATACACTTCTCT 2129
777 GluLysAspAlaIleLysLeuAlaGlyLys.....LeuLysAsnLeuLys 792
2130 .....GGCTATATCCGCTTCATTTGGCAATATGAAATCTGCACAAC 2173
792 smetCysLeuPheHisLeuThrHisLeuSerAsp.....IleGlyG 806
2174 TCTG.....TTTCTCAGATGATACGATCTCATTTGGGG 2205
806 luGlyMetAspTyrIleValLysSerLeuSerSerGlu..... 818
2206 AAATTCCTTCAATTTGTGTGCAATTTGACATCATCTGAAGTGTGTATATG 2235
819 .....ProCys.....AspLeuG1 823
2256 TCGAGAAACAATTTGAAGGAAAGCTCCGCAATGTTGGGTAATATACG 2305
823 uGluIleGlnLeuValSerCysLeuSerAlaAsnAlaValLys...I 839
2306 TGACCTTCACATTTTGTGCG.....ATGTCATCTAATAGTTTCAGAGAG 2349
839 leLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIleLeuAspLeu 855
2350 AGCTCCCTTCAATCTAATTTCAATTAACATCATCAAAAATCTGATTTT 2399
856 serGluAsnTyrLeuGluLysAsp.....GlyAsnGluAl 867
2400 GCGAGAAACAATCTGAGAGGAGCAATACCAATTTTGGCAATATTAG 2449
867 aLeuHisGluLeuIleAspArgMetAsnValLeuGlnLeuThrAlaL 884
2450 TAGCTTCAGGTTTGTGATATGAGAAAT.....AACAACTTCTGGG. 2492
884 euMetLeuProTTrpGlyCysAspValGlnGlySerLeuSerLeuLeu 900
```

```

2493 ..ACTCTTCCAACAAAATTTAGCATTTGATGTTCACTGATAGTCTCAAC 2540
901 LysHis.....LeuGluGluValPro..... 907
2541 TTGCATTCGCAATGACACTAGACAGATGAAATCCCTCGCTTTGGACAATTG 2590
908 ...GlnLeuValLysLeuGlyLeuLysAsnTrpArgLeuThr 922
2591 CAAAAGCTGCAGATGTTGATTTAGAGACAAATCACTCAACGACACAT 2640
923 .....GluLeuArgIleLeuGlyAlaPhe 930
2641 TTCCCATGTGTTGGAACTTTGCCAGAGCTGAGATTTTAAGTTTGACA 2690
931 PheGlyLys.....AsnProLeuLys..... 937
2691 TCGAATTAATTTGCATGACCTTAAGATCATCAGGCGCTGAATATGTT 2740
938 ...AsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAspG 953
2741 TCCTGATCTCGAATCATAGATCTCTCTCGCATGCTATTCGCAAGAC. 2789
953 LysTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLeuValPhe 969
2790 .....TTACCAACGAGCTCTATTGAAACATTTGAAGGATGAGACA 2831
970 PheAspPheSerThrLysGlu.....PheLeuProAspPr 981
2832 GTTGATTAACAAATGAGAGCAACCAAGTTATGAAAGCTATTACGATGACTC 2881
981 AlaLeuVal.....ArgLysLeuSerGlnValLeuS 992
2882 GGTGCTATTGTGCACAAAGGAGTTGGAGCTTGAATTTGTGACAATCTCT 2931
992 eLysLeuThrPheLeu 997
2932 CTTTGTACACAAATTATTC 2948

```

seq_name: /cgrn2_6/prodata/2/lna/5b_COMB.seq:US-09-180-439-2

```

seq_documentation_block:
; Sequence 2, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatzixanthis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; EARLIER FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; EARLIER FILING DATE: 1996-09-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3979
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
; US-09-180-439-2

```

```

alignment_scores:
  Quality: 172.00      Length: 639
  Ratio: 0.533        Gaps: 38
  Percent Similarity: 50.348    Percent Identity: 23.005
alignment_block:

```

US-09-697-089-2 x US-09-180-439-2 ..
Align seg 1/1 to: US-09-180-439-2 from: 1 to: 3979

```

459 LeuThrSerHisGluProGlnGluValThrLysGlyAsnGlyTyrLeuG 475
1260 CTTTCGGCTTTATTCCTGACAGAAATA.....GGTTACTTAG 1297
475 nLysMetValSerIleSer..AspIleThrSerThrTyrSerSerLeu 491
1298 GTCTCTTACTAGCTATCTTGATATCAAC..... 1328
491 euArgTyrThrCysGlySerSerValGluAlaThrArgAlaValMetLys 507
1329 .....TTTCTTAGTGTTCCATTCCGCTTCATTTGGGGAATGTGAAC 1370
508 HisLeuAlaAlaValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAl 524
1371 AACTGTCTTTTGTATCTTACATATATACAGCTTTCTGGCTTATTT.. 1418
524 aLysArgProLeuThrPArgGlnGluSerLeuGlnSerValLysAsnThr 541
1419 .....CTGAAAGAAATAGCTTACCTAAGGTCACCTTA 1449
541 hrGluGlnGluIleLeuLysAlaIleAsnIleAsnSerPheValGluCys 557
1450 CTAAAG.....CTATCTTTG 1463
558 GlyIleHisLeuTyrGlnGluSerThrSerLysSerAla..... 570
1464 GGTATCAACTTTCTTAGTGTTCCATTCCTGCTTATTTGGGGAATCTAA 1513
571 ...LeuSerGlnGluPheGlnAlaPhePheGlnGlySerLeuTyrI 586
1514 CAACCTGTCTAG.....TTGATCTTTTACA 1539
586 Leasn.....SerGlyAsnIleProasp.....TyrLeu..... 595
1540 ATAATAGCTTTCTGCTCTATCTCTGAAAGAAATAGTTACTTAAGGCT 1589
596 .....PheAspPhePheGluHis.....LeuProAs 604
1590 CTACTTACTAGATTGGTGAGAGATGCTCTTATGCTCTATTCCTTC 1639
604 nCysAlaSerAlaLeuAspPheIle..LysLeuAspPheTyrGlyGly 620
1640 TTCATTGGGGAATCTTAACACTGTCTAGTTGGATCTTTACAAATAATA 1689
620 IametalSerTrpGluLysAlaIleGluAspThrGlyGlyIleHisMet 636
1690 AGCTTTCTGGC.....TCTATCTCGAAGAAATAGTTACTTAAGGCTCT 1733
637 .....GlnGluAlaProGlnThrTyrIleProse 646
1734 CTACTTACTAGATTGGTGAGAGATGCTCTTATGCTCTATTCCTG 1783
646 rArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThrLeug 663
1784 TTCATTGGGGAATCTG.....AACACTTGTTATGATGCT 1818
663 LuValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyr 679
1819 ATCTTTACATATATCAGCTTTCTGGCTCTATTCCTGAAAGAAATAGCTTAC 1868
680 LeuGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysAr 696
1869 CTG.....AGTTCTCTTACTGACACTATATTTGGGGAATAAAC.. 1904
696 gCysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysA 713
1905 .....TCTCTTAATGCTCTATTCCTGCTTCAATTTGGGGAATCGAAC 1947
713 snIleTyrSerLeuMetValGluAlaSerProLeuThr.....Ile 726

```

```

1948 ACTGTGTTATGTTGATCTTTACATATATACGCTTTCTGCGCTATTTCT 1997
727 GUAAPGCUAARGHISILETHRSERVALTHRASNLEULYSTRLEUSERIL 743
1998 GAGAAATAGGTACCTGAGTCTCTTACTGAACTA..... 2033
743 eHISAPLEUgInASngInARgLEUProGlyGlyLEUThASpSERLeUG 760
2034 ....TTTTGGTAATAACTCTCTTAATGGCTCTATTTCTGCTTCATTTG 2079
760 lYASnLEULySAsnLEUThrLYSLEULlEMELASpASnILEYsMetASn 776
2080 GGAATCTAAACAACCTGTCTAGCTTGATCTTTACATATATACAGCTTCT 2129
777 GUGLUASpAlAlELySLEULAGLUgLY...LEULySAsnLEULySLY 792
2130 ....GGCTCATTCCTGCTCATTTGGCAATATAGAAATCTGCAAC 2173
792 sMetCysLEUThrHISLEUThrHISLEUSERAsp.....lEGLYG 806
2174 TCTG.....TTTTCACTGATTAACGATCTCATTTGGGG 2205
806 lUGlYMeLAspTYRILEVALySSErLEUSERSerGU..... 818
2206 AAATTCCTTATTTGTGTGCAATTTGACATCACTGCAAGCTGTGTATAG 2255
819 .....ProCys.....AspLEUGl 823
2256 TCGAGAAACAATTTGAAGGAAAGTTCCGCAATGTTGGTAATATCAG 2305
823 uGUlIEGInLEUVALSerCysLEUSERAlASnAlVALyS...I 839
2306 TGACCTTCACATTTTGTG....ATGTCATCTAATAATGTTTCAGAGGAG 2349
839 lEULyAGInASnLEUnIASnLEUVALySLEUSERILEULySLEU 855
2350 AGTCCCTTATCTATTTCCAATTTAACATCACTAAATACTTGATTT 2399
856 SerGLuASnTYRLEUGLUySAsp.....GLyASnGLUAl 867
2400 GCGAGAAACAATCTGGAGGAGCAATACACAATTTTGGCAATATTAG 2449
867 AlEUnHISGLULEULeASpARGMeLAsnVALLEUGLUInLEUThrAl 884
2450 TACCTTCACAGTTTGTGATTCGAGAT.....AACAAACTTTCCTGGG 2492
884 eUeTLEUProTYRGLYcYAspVALGInGLYSErLEUSERSerLEU 900
2493 ..ACTCTCCACAACAATTTAGCATTTGATGATCTGATTAAGTCTCAAC 2540
901 LysHIS.....LEUGLUInVALPro..... 907
2541 TTCATTCGCAATGACATAGCAATGAATCCCTGCGCTTTGGACAATTG 2590
908 ....GInLEUVALySLEUGLYLEUySAsnTYRPARGLEUThrASpHR 922
2591 CAAAAGCTCAAGTCTTGAATTTAGGAGACAATCAACGACACAT 2640
923 .....GluLEARgILELEUGLYAlAPhe 930
2641 TTCCCATGTGTTGGAACTTTGCCAGAGCTGAGAGTTTAAGGTGACA 2690
931 PheGLYLYS.....AsnProLEULyS..... 937
2691 TCGAATAAATTCGATGACCTATAGATCATCAAGGGCTGAATCATGTT 2740
938 ....AsnPhEGLInLEUASnLEUAlAGLYAsnARGVALSerSErASpG 953
2741 TCTGTATCTCGAATCATGATCTCTCCCAACGATTCCTCGCAACAC 2789
953 lYTRPLeUAlAPheMetGLYVALPheGLUASnLEULySGLInLEUVALPhe 969

```

```

2790 .....TTACCAAGAGTCTATTGAACATTTGAAGGAGTACAGACA 2831
970 PheAspPheSerThrLYSGlu.....PheLEUProAspPr 981
2832 GTTGATTAACAATGAGAGCAACAGTTATGAAGGATATTACGATGACTC 2881
981 oAlALEUVAL.....ArgLYSLEUSERGLNVALyS 992
2882 GGTGGTAGTTGTGACAAAGAGATTGAGCTTGAAATTTGTGAGAAATTCGT 2931
992 eRLySLEUThrPheLEU 997
2932 CTTTGTACACATTTATC 2948

seq_name: /cgn2_6/ptodata/2/lna/6B.COMB.seq:US-09-353-585-4

seq_documentation_block:
; Sequence 4, Application US/09353585
; Patent No. 6287865
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; Jones, David A
; Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses
; thereof
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nixon & Vanderhye PC
; STREET: 8th Floor, 1100 No. 6287865th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/353,585
; FILING DATE: 15-Jul-1999
; CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q
; 1/68
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/930,277
; FILING DATE: 27-Oct-1997
; APPLICATION NUMBER: PCT/GB96/00785
; FILING DATE: 01-Apr-1996
; APPLICATION NUMBER: GB 9506658.5
; FILING DATE: 31-Mar-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-69
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tomato
; STRAIN: Cf2
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-353-585-4

```

```
alignment_scores:
  Quality: 165.00      Length: 645
  Ratio: 0.546         Gaps: 31
Percent Similarity: 46.822  Percent Identity: 21.395
```

```
alignment_block:
US-09-697-089-2 x US-09-353-585-4
```

Align seg 1/1 to: US-09-353-585-4 from: 1 to: 3573

[illegible]

664	lThrlleuAaGAPhPheSerLysLysAsnLysGlnAspIleThrTyrLeuG	681
665	
937	CGAGCTTTGAGAGACAAGCTCTTAAT	964
681	LylsIlePheSerSerAlaThrSerLeuAArgLeuGlnIleLysArgCys	697
964	964
698	AlaGlyAlaAlaGlyLeuSerLeuValIleuSerThrCysLysAsnIl	714
965GGCCTATTCTCCGTCCTTCATTATGGGGAATCTGAAAACCTT	1002
714	eTyrSerLeuMetValGlnAlaSerProLeuThrIleGlnAspGlnArgH	731
1003	GTCAGAGTTGAATCTTGTAATAATACAGCTTCTCGCTCTATCTCT	1048
731	IsIleThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGln	747
1049GCTTCATTTGGGGAATCTGCAACACTTGTATGTCATGTCATTAC	1093
748	AsnGlnArgLeuProGlyLysLeuThrAspSerLeuGlnAsnLeuLysAs	764
1094	AAATACACACTTTCGGCTCTATTCCTCGCTTCATTGGGGAATCTGAACA	1143
764	nLeuThrLysLeuIleMetAspAsnIleLysMetAsnGlnLysAlaI	781
1144	CTGTCTATGTTGTATCTTACAAATATACAGCTTCTCGCTCTATCTCGT	1193
781	IeLysLeuAlaGlnLysLeuLysAsnLeuLysLysMetCysLeuPheHis	797
1194	CTTCATTTGGGG...AATCGAACAACCTTGTACAGCTGTATCTACAT	1240
798	LeuThrHisLeuSerAspIleGlyGlnLysMetAspTyrIleValLys	813
1241	AATACACTTTCGGCTCTATTCCTGAGAAATAGTTACTTGAGTTCTCT	1290
814SerLeuSerSerGluProCysAsp	821
1291	TACTTATCTAGATTTGACAAATAACTGCATTAATGATTTATTCCTGCTT	1340
822LeuGlnGlnIleGln	826
1341	CATTGGCAATATGACAACTGCTTTTTGTTGTTCTTATTAAGAAATCAG	1390
827	LeuValSerCysCysLeuSerAlaAsnAlaValLysIleLeuAlaGlnAs	843
1391	CTTGCTAGCTCT...GTTCCGAGAGA	1413
843	nLeuHisAsnLeuValLysLeuSerIleLeuAspLeuSerLysAsnTyrL	860
1414	AATAGGTTACCTAAGGCTCTTAATGCTGCTGATTTGAGTACAGATGCTC	1463
860	eGlnLysAspGlnAsnGlnAlaLeuHisGlnLeuIleAspArgMetAsn	876
1464	TT...AATGCTCTATTCCTGCTTCAATGGGAATTTGAC	1501
877	ValLeuGlnGlnLeuThrAlaLeuMetLeuProTrrGlyCysAspValG1	893
1502	AACCTGTAGGTGAATCTT...GTTAA	1527
893	nGlySerLeuSerSerLeuLeuLysHisLeuGlnGlnValProGlnIle	910
1528	TAAATCAGCTTTCGGCTCTAATCTT...GAAATAATAGGTTAACTTAA	1571
910	AlLysLeuGlyLeuLysAsnTrrArgLeuThrAspThrGlnIleArg...	925
1572	GGCTCTTAATGCTCTGAT...TTGAGTACAGAAATGCTTAATAGGC	1615
926	IleLeuGlyAlaIlePheGlyLysAsnProLeuLysAsnPheGlnGlnIle	942
1616	TCTATTCCTGCTTCAATGGGGAAT...TTGAACAACCTGTCTAGTGT	1659

```

942 uasnleuAlaGlyAsnArgValSerSerASP.....GlyT 954
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1660 GAATCTGTTAAATATCATCAGCTTCTGCTATTCCTGAGAAATAGGT 1709
954 rpleu..... 955
1710 ACCTAAGATCTCTTAATGACCTAGCTTGAAGTGAATGCTCTTAATGCC 1759
956 .....AlaphmetGlyValPheGluAsnLeuLysGlnLeuValPh 969
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1760 TCATATCTGCTTCATTCGGGAATCTGAGAACATCTGCTATGATGTAATCT 1809
969 ePheAsPheSerThrLysGlnPheLeuProAsPProAlaLeuValArgL 986
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1810 TTACAATAATCATCAGCTTCTGCTCTATTCCTGAGAAATA..... 1849
986 ylsLeuSerGlnValLeuSerLysLeuThrPheLeu 997
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1850 .....GGTTACTTGAAGTTCTCTTACTTATCTA 1876
name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-353-585-1

```

documentation block:

```

Sequence 1, Application US/09353585
Patent No. 6287865

```

```

GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
Jones, David A

```

```

TITLE OF INVENTION: Plant pathogen resistance genes and uses
Jones, Jonathan DG

```

```

NUMBER OF SEQUENCES: 15
theoret

```

```

CORRESPONDENCE ADDRESS:
ADDRESSER: Nixon & Vanderhye PC

```

```

STREET: 8th Floor, 1100 No. 6287865th Glebe Road
CITY: Arlington

```

```

STATE: Virginia
COUNTRY: United States of America

```

```

ZIP: 22201-4714

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585

```

```

FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q
1/68

```

PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997

```

```

APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996

```

```

APPLICATION NUMBER: GB 950658.5
FILING DATE: 31-MAR-1995

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson

```

```

REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000

```

```

TELEFAX: (703) 816-4100

```

```

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```

```

LENGTH: 6471 base pairs
TYPE: nucleic acid

```

```

STRANDEDNESS: double
TOPOLOGY: linear

```

```

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```

```

ORIGINAL SOURCE:
ORGANISM: tomato

```

```

; STRAIN: Cf2
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1754..5012
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1677..1753
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-353-585-1

```

alignment_scores:

```

Quality: 165.00 Length: 645
Ratio: 0.546 Gaps: 31
Percent Similarity: 46.822 Percent Identity: 21.395

```

alignment block:

```

US-09-697-089-2 x US-09-353-585-1 ..

```

```

Align seg 1/1 to: US-09-353-585-1 from: 1 to: 6471

```

```

425 GlyLeuLeuCysLysTyrThrAlaGlnArgPheLysProLysTyrLysPh 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2094 GGTTCCTAGCCCAAGCTTCACATATCCGC.....AT 2125

```

```

441 ePheHisLysSerPheGlnGluTyrThrAlaGlyArgLeuSerL 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2126 ATTTCACAAATCA..... 2138

```

```

458 euleuthSerHisGluProGlnGluValThrLysGlyAsnGlyTyrLeu 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2139 ..TTAAATGAGATTATTCCTTAAGAAATA.....GGTTACTTA 2174

```

```

475 GlnLysMetValSerIleSerAspIleThrSerThrTyrSerSerLeu 491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2175 AGGTCTCTTACTTAAGCTATCT..TTGGGTATCAACTTTCTTAGT.... 2216

```

```

491 uArgTyrThrCysGlySerSerValGluAlaThrArgAlaValMetLys 508
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2217 .....GGTTCATTCCTGCTTCAGTGGGAATCGAATCA 2250

```

```

508 IseuAlaAlaValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAla 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2251 ACTTGCTCTTTTGTATCTTCAATAATACGCTTTCGGCTCATTT... 2297

```

```

525 LysArgProLeuThrPargGlnGluSerLeuGlnSerValLysAsnThr 541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2298 .....CCTGAAGAAATTAAGTTAAGATCTCTTAC 2329

```

```

541 rGlnGlnGluIleLeuLysAlaIleAsnIleAsnSerPheValGluCys 558
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2330 TGACCTAGATTGG..... 2342

```

```

558 LylIeHisLeuTyrGlnGluSerThrSerLysSerAlaLeuSerGln 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2343 .....AGTGATTAATGCTTAAATGGCTCT 2366

```

```

575 PheGluAla.....PhePheG1 580
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2367 ATTCTGCTTCATTCGGGAATATGAGAACTTGCTTTTGTTCCTTTCTTA 2416

```

```

580 nGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsP.....T 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2417 TGGAAATCAGCTT.....TCGGCTCATTCCTGAAGAAATATATGT 2457

```

```

594 yrlLeu.....PheAsPhePheGlnHisLeuProAsnLys 605
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2458 ACCTAAGATCTCTTACTTACCTAGATTGAGTGAAGAACTCTTAATGGC 2507

```

```

606 Ala.....SerAlaLeuAspPheIleLysLe 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2508 TCATATCTGCTTCATTCGGGAATTTGAACAACCTTGCTTTTGTTCCT 2557

```



```

614 uasphetyrglyAlaMetAlaSerTrpGluValAlaGluAsp 631
      |||||: : : : : |||||:
2558 T.....TATGGAATCAGCTTCTGCG.....TCTATTCCGAGAGAA 2595
      |||||: : : : : |||||:
631 hrclyglyIleHisMetGluGluAlaProGluThrIleProSerArg 647
      |||||: : : : : |||||:
2596 TAGCT.....TACCTA..... 2606
      |||||: : : : : |||||:
648 AlaValSerLeuPheAsnTrpLysGlnGluPheArgThrIleVal 664
      |||||: : : : : |||||:
2607 .....AGACTCTTAATGT 2620
      |||||: : : : : |||||:
664 LThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuG 681
      |||||: : : : : |||||:
2621 CCTAGCTTTGAGTGAGATCCCTTAAT..... 2648
      |||||: : : : : |||||:
681 LysLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArgCys 697
      |||||: : : : : |||||:
2648 ..... 2648
      |||||: : : : : |||||:
698 AlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysAsnI 714
      |||||: : : : : |||||:
2649 .....GGCTTAATCCTCGCTTCATTTGGGGAATCTGAAGAACTT 2686
      |||||: : : : : |||||:
714 eTySerLeuMetValGluAlaSerProLeuThrIleGluAspGluArgH 731
      |||||: : : : : |||||:
2687 GTCTAGGTTGAATCTTGTAATAATACAGCTTTCGGCTCTATTCTT... 2732
      |||||: : : : : |||||:
731 LsIleThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGln 747
      |||||: : : : : |||||:
2733 .....GCTTCATTTGGGGAATCTGAACACTGTCATGTGTGTTACTTTC 2777
      |||||: : : : : |||||:
748 AsnGlnArgLeuProGlyLeuThrAspSerLeuGlyAsnLeuLysAs 764
      |||||: : : : : |||||:
2778 AAATACCACTTTCTGCTCTATTCTCGCTTCATTTGGGGAATCTGAACA 2827
      |||||: : : : : |||||:
764 nLeuThrLysLeuIleMetAspAsnIleLysMetAsnGluGluAspAla 781
      |||||: : : : : |||||:
2828 CTTCGTCTATGTTCTATCTTACATAATACAGCTTTCGGCTCTATTCTCG 2877
      |||||: : : : : |||||:
781 LelLysLeuAlaGlyLysLeuLysAsnLeuLysLysMetCysLeuPheHis 797
      |||||: : : : : |||||:
2878 CTTCATTTGGGG...AATCTGAACAACTGTCATGTGTTACTCTACAAAT 2924
      |||||: : : : : |||||:
798 LeuThrHisLeuSerAspIleGlyGluGlyMetAspTyrIleValLys.. 813
      |||||: : : : : |||||:
2925 AATCACCTTTCTGCTCTATTCTCGTGAAGAAATAGTTACTTGAGTTCTCT 2974
      |||||: : : : : |||||:
814 .....SerLeuSerSerGluProCysAsp..... 821
      |||||: : : : : |||||:
2975 TACTTATCTAGATTGAGTAATACTCCATTAATGATTTATTCCTGCTT 3024
      |||||: : : : : |||||:
822 .....LeuGlnGluIleGln 826
      |||||: : : : : |||||:
3025 CATTTGGCAATATGAGCAACTTGGCTTTTGTGTTCTTATGAAATCAG 3074
      |||||: : : : : |||||:
827 LeuValSerCysCysLeuSerAlaAsnAlaValLysIleLeuAlaGlns 843
      |||||: : : : : |||||:
3075 CTTCGTAAGCTT.....GTTCCTGAAGA 3097
      |||||: : : : : |||||:
843 nLeuHisAsnLeuValLysLeuSerIleLeuAspLeuSerGluAspTyrL 860
      |||||: : : : : |||||:
3098 AAATAGSTTACSTAGGCTCTTAAATCTCTGATTTGAGTAGAAGAACTGC 3147
      |||||: : : : : |||||:
860 euGlnLysAspGlyAsnGluAlaLeuHisGluLeuIleAspArgMetAsn 876
      |||||: : : : : |||||:
3148 TT.....AATGCTCTATATTCCTTCATTCGGAATTTGAAAC 3185
      |||||: : : : : |||||:
877 ValLeuGlnGluLeuThrAlaLeuMetLeuProTrpGlyCysAspArgAla 893
      |||||: : : : : |||||:
3186 AACTGTCTTAGGTTGAACTT.....GTTA 3211
      |||||: : : : : |||||:
893 nGlySerLeuSerSerLeuLeuLysHisLeuGlnGluValProGlnLeu 910

```

```

.....: : : : : |||||:
3212 TATACGCTTCTGGCTATTCCT.....GAGAAATAGGTTACCTTA 3255
      |||||: : : : : |||||:
910 allLysLeuGlyLeuLysAsnTrpArgLeuThrAspThrIleuIleArg... 925
      |||||: : : : : |||||:
3256 GGTCTCTTAATGCTCTGAT.....TTGAGTGAGAAATGCTCTTAATGCG 3299
      |||||: : : : : |||||:
926 IleLeuGlyAlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLe 942
      |||||: : : : : |||||:
3300 TCTATTCCGCTTCATTCGGGAAAT.....TTGAACAACTGTCTAGGTT 3343
      |||||: : : : : |||||:
942 uAsnLeuAlaGlyAsnArgValSerSerAsp.....GlyT 954
      |||||: : : : : |||||:
3344 GAATCTGTTTAATAATACACTTTCGCTCTATTCCTGAGAAATAGGTT 3393
      |||||: : : : : |||||:
954 rPLeu..... 955
      |||||: : : : : |||||:
3394 ACCTAAGATCTCTTAATGACCTAGGTTTGAGTGAGAAATGCTCTTAATGCG 3443
      |||||: : : : : |||||:
956 .....AlaPheMetGlyValPheGluAsnLeuLysGlnLeuValPhe 969
      |||||: : : : : |||||:
3444 TCTATTCTGCTTCATTTGGGGAATCTGAACAACTGTCTATGTTGATCT 3493
      |||||: : : : : |||||:
969 ePheAspPheSerThrLysGluPheLeuProAspProAlaLeuValArgL 986
      |||||: : : : : |||||:
3494 TTACAAATATATCAGCTTTCGGCTCTATTCTCTGAGAAATA..... 3533
      |||||: : : : : |||||:
986 yLysLeuSerGlnValLeuSerLysLeuThrPheLeu 997
      |||||: : : : : |||||:
3534 .....GGTTACTTGAAGTTCTCTTACTATCTA 3560

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-180-439-7

```

seq_documentation_block:
: Sequence 7, Application US/09180439
: Patent No. 6225532
: GENERAL INFORMATION:
: APPLICANT: Dixon, Mark S
: APPLICANT: Hatziathanis, Kostas
: APPLICANT: Jones, David A
: APPLICANT: Jonathan DG
: TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
: FILE REFERENCE: 620 - 53
: CURRENT APPLICATION NUMBER: US/09/180,439
: CURRENT FILING DATE: 1998-12-06
: EARLIER APPLICATION NUMBER: PCT/GB97/01249
: EARLIER FILING DATE: 1997-05-08
: EARLIER APPLICATION NUMBER: GB 9609681.3
: EARLIER FILING DATE: 1996-05-09
: EARLIER APPLICATION NUMBER: GB 9619924.5
: EARLIER FILING DATE: 1996-09-24
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7
: LENGTH: 4123
: TYPE: DNA
: ORGANISM: Lycopersicon esculentum
US-09-180-439-7

```

```

alignment_scores:
  Quality: 164.00      Length: 666
  Ratio: 0.488        Gaps: 38
  Percent Similarity: 50.450    Percent Identity: 22.072

```

alignment_block:

US-09-697-089-2 x US-09-180-439-7 ..

Align seg 1/1 to: US-09-180-439-7 from: 1 to: 4123

```

459 LeuThrSerHisGluProGlnGluValThrLysGlyAsnGlyTyrLeuG 475
      |||||: : : : : |||||:
1260 CTTCCTGCGCTTATTCCTGAAGAAATA.....GGTTACCTAAG 1297

```



```

2945 TCTATTGAACTTTGAAAGGATGAGACAGTTGATTAACAATGAGAG 2994
967 lu.....PheLeuProAspProAlaLeuVal..... 984
2995 AACCAAGTATGAAAGCTATTACGATGACCTCGGTGAGTGTGACAAAG 3044
985 .....ArgLysLeuSerGlnValLeuSerLysLeuThrPheLeu 997
3045 GGAATTGAGCTTGAAATGTGACAAATTCCTCTTTGTACCAATTTATC 3092
seq_name: /c9n2_6/plodata/2/1na/5A_COMB.seq:us-08-353-700-2

```

```

seq_documentation_block:
; Sequence 2, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOEFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
;
us-08-353-700-2

```

```

alignment_scores:
  Quality: 149.50      Length: 1156
  Ratio: 0.266        Gaps: 52
  Percent Similarity: 48.702  Percent Identity: 18.426

```

alignment_block:

```

us-09-697-089-2 x us-08-353-700-2 ..
Align seg 1/1 to: us-08-353-700-2 from: 1 to: 10136

```

```

39 AsnIleIleCysGluLysValGlnAspAlaIleArgGlyIleI 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2220 AACGTGTAGACAGTAAGTCAGTACGAGGAGTAAGACCCAGAAA...CTAGC 2266

```

```

55 eHisMetIleLeuLysGlySerGluSerCys..... 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2267 TTAATGAGAGCTACAGCAAAAGCTAGTTCAGATCGAAACATCAGA 2316
67 .....AsnLeuPheLeuLysSerLeuLysGluTrpAsnTyrPro 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2317 AGGAATAGAAAATATGCTTTGAAAGACT..... 2345
80 LeuPheGlnAspLeuAsnGlyGln.....SerLeuPheHis.....G 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2346 .....TCTACGCTTACTGCGCAAGTTGAAGATCTGAAGACACAGCTTGA 2389
92 nThrSerGluGlyAspLeuAspLeuAlaGlnAspLeuLysAspLeu 109
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2390 GTTACTGTCAAAATGAAATATGACACAAAGCCGGTGTACCAAGCTTGC 2439
109 yHisThrProSerPheLeuAsnPheTyrProLeuGlyGluAspLeu 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2440 ATGCCGATATGACAGCCTCAGAGATCTGCTAAATCCAAAGATGCTTCT 2489
126 IleIlePheAsn.....LeuLysSerThrPheThrGln 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2490 GTGGTACAAATGAAGATCATCAGACAGAGCTTTGGCTTTGATCAGCA 2539
136 uProValLeuTrpArgLysAspGlnHisHisArgValGluGlnLeu 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2540 GCCTGCCATG.....CATCATTCCTTTGCAAAATATATATG 2574
153 hLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleIleGly 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2575 GAGACAAAGCAAGCATGCTTCAGAGAGGATGATGCTGTTAAGCA 2624
170 GluSerGlyLysGlyLysSerThrLeuGln...ArgIleAlaMetI 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2625 GACCAAGTCCGAAAATTTGCCATCTCAAAATAGAGTTGATTCAT 2674
185 uTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPheValPhe 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2675 T.....GAATTT 2682
202 hLeuArgLeuSerArgAlaGlnGlyLysLeuPheGlnThrLeuCysAsp 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2683 CATTAGAGTCTCAAAAACAGATGATGACTGACCTCGAAAGCAGTGTGA 2732
219 GlnLeuLeuAspLeuProGlyThrIleArgLysGlnThrPheMetAla 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2733 GAGTTGGTCCAATCAAAAGAGAAATAGAGAAAT..... 2768
235 tLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTyrAsn 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2769 .CTCATGAAGCAGACAGCATG.....CATCAAA 2796
252 hLeuPheLysProGlnAsnCysProGlnIleGlnAlaLeuIleLysGlu 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2797 GTTTGTGCTGTAACAAAGTCAGCGCATTAAGTTACAGAGACAGACT 2846
269 HisArgPheLysAsnMetValIleValIleThrThrThrGluCysLeu 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2847 TCTGCTCACCAGAAAT...GTGTGCTGAACCTTAAGTCCCTTGAGAA 2893
285 gHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMetThr 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2894 CAAGCAAAAAGAGCTGCACACTTTAAATATATAGCTA..... 2930
302 hAspSerAlaGlnAlaLeuIleArgGluVal...LeuIleLysGluLeu 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2931 ..GAAACTGACGACGACAGATTCAAAGATTAAAAAAGCACAACATCTA 2978
318 AlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsn 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||
2979 CTGTGAAGACTCTCTAAAGAGGCTTACAACTTTATCCGAAACCTTAAGCT 3028
334 uMetLysThrProLeuPheValValIleThrCysAlaIleGlnMetGly 351

```

```

3029 GGAGAGCAAGAAATGAGTTCATCATTTCT.....TTAAATA 3066
351 luserGluPheHisSerHisThrInThr.....ThrLeuPheHisThr 365
3067 AAAGGAAATTTGAAGAGCTGACCCAGAGAAATGGACTCTTAAGAAATT 3116
366 PheTyrAspLeu.....LeuIleGlnLysAsnLys.. 375
3117 AATGATCCTTAATCAAGAGAGTGAATCAATCCAGAAAGCTGACAG 3166
376 .....HisLysHisLysGlyAlaAla...AlaSerA 385
3167 TTTTGCAAACTATATAGATGAAGGAGAAAGCATTTTCAGAGTTATCTG 3216
385 sPpheIleArgSer.....LeuAspHisCysGlyAspLeu 396
3217 ATCAGTACACCAAGAAAACTTATTTACTACAAAGATGTGAAGAAACC 3266
397 .....AlaIleuGlnGlyAlaPheSerHisLysPheAspPheGlnLeuG1 411
3267 GGAATGCAATGATGAGATCTT...AGTCAAAATATACAA...GCAGCACA 3310
411 nAspValSerSerValAsnGluAspValLeuLeuThrThrGlyLeuLeu 428
3311 GGAAGAAGATTTCTAAATTAGATGCTGCTTAATGCAATGCACTAGTCTTT 3360
428 yslLysTyrThrAlaGlnArgPheLysProLysTyrLysPhePheHisLys 444
3361 GTGAATAATGCAAAATGATGTTGGACACGCTAAAGAGCAATTTTGCAAG 3410
445 SerPheGlnGlyTyr.....ThrAlaGlyArgArgLeuSerSerLeu 449
3411 GAACACCAAGATTTCTTAACAAATATGATCTGCTGAAGAAAGAAATCA 3460
450 .....ThrAlaGlyArgArgLeuSerSerLeu 459
3461 GAATCTGATGCTAGAGTTGAGACAGCTGACAGCTCTGAGATCTGAGA 3510
459 eutPheSerHisGluProGlnGluValThrLysGlyAsnGlyTyrLeuGln 475
3511 TGACAGATTAACCAAAACAATCTTAAGACGAGCGTGTGTTTAAAGCAA 3560
476 LysMetValSerLysSerAsp.....IleThrSerThrTyrSer 488
3561 GAAATCAATGATTTAAAGGAACCAAAATGCAAAAGGAAGTAA 3610
488 rSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThrArgAla 505
3611 TGACTTATTACAGAGAT.....GACACGCTGATGAAGG 3645
505 alMetLysHisLeuAlaAlaValTyrGlnHisGlyLysLeuLeuGlyLeu 521
3646 TAATGAG.....ACTAACAATGATGT..... 3668
522 SerIleAlaLysArgProLeuTyrArgGlnLysLeuGlnSer.... 536
3669 .....CAAATCTAGAAATGAGAAC 3688
537 ValLysAsnThrThrGlnGlnGlnIleLeuLysAlaIleAsnIleAsn 553
3689 AATTAGGAACCTGTGAAAGAAAGAGAGAGTGAAGAAATCAATGTAATT 3738
553 ePheValGlyCysGlyIleHisLeuTyrGlnLysSer...ThrSerLys 568
3739 TTTAAACCTCAATGATGCTTGAAGTTAAAGAAATTTCTTAGATAGTTAT 3788
569 SerAlaLeuSerGlnGluPheGlnAlaPhePheGlnGlyLysSerLeuTyr 585
3789 AATCGCGAGTGTGTCATAGAACCTATGCTTAAGAAATTAAGAAATTA 3838
585 rIleAsnSerLysAsnIleProAspTyrLeuPheAspPhePheGlnHisL 602

```

```

3839 ACTTCAGGAAGT.....GAGAGG 3858
602 eutProAsnCysAlaSer...AlaLeuAspPheIleLysLeuAspPheTyr 617
3859 AGAAGAGACTGCTGCACAGCATTAATTACAGCAATTAAGAGAGATCTTGA 3908
618 GlyLysAlaMetAlaSerTyrGlnLysAlaAlaGlnAspThrGlyGly 634
3909 ACCAGCAATTTGCAAGACATGACAG.....TCACAAAGAAATTAAGTGCCT 3952
634 eHisMetGlnGlu.....AlaProGlnThrTyrIleProSer...Arg 648
3953 TTAAGACTGTGAATATAGATGCGAAGAAAGATATATTCAGGGCTCATG 4002
648 lAlaIleSerLeuPhePheAsnTyrPlyGlnGluPheArg...ThrLeuGln 663
4003 AGTGTCTACACAGTCAAAACGACAAATGACACCTTCAGTGTCTGTGCA 4052
664 ValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrLe 680
4053 ACAACAATGACACAGCTGAATGAGCTAGAGAAATATGTGAATATCTGCA 4102
680 uGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArgC 697
4103 GGTGAAAGATATGAACTGTAACTGAGCTGAAGTAAATGATCAAGTCGAAT 4152
697 yslAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLysAsn 713
4153 CTATCACACGACATAGAAATGCGACAGAGGTAGGAAACTCTAAT 4202
714 IleTyrSerLeuMetValGluAlaSerProLeu.....ThrI1 726
4203 GAAGTTAAATATTAATATGATGACAGCTGCTCTCATGGTGAATAGT 4252
726 eGluAsp.....GluArgHis 732
4253 GGAAGACATACCAGAGGTGAATTTGTTGAACAACAATGAACAGCAC 4302
732 lPheSerValThrAsnLeu.....LysThrLeuSer 742
4303 CTGTGTCTTTGGTCCATCTGACAGAGATATTCCTACGACACTTGCA 4352
743 IleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
4353 TTGTCAGACAAAGAAATTCAAATGACACTTTCGCAATTTGCAAGAAATT 4402
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAsp..... 771
4403 CTATCTTTACAAAGTGAACACAAATTTTACATGATCAGCAGCTGCA 4452
771 ..... 771
4453 TGAGCTTAATATGTCAGAGCTGCAGACCTATGTTGACTATTAAAGCC 4502
772 .....AsnIleLysMetAsnGlnGluAspAlaI1 781
4503 GAAATTTTGTCTTGTCAACCAATCTGAGAACTTCAAGGTGACTGTGT 4552
781 eLys.....LeuAlaGlnGlyLeu...LysAsnLeuLysL 792
4553 GAAGGAGATGAGCTGGCTTGGAGAGGGCTGCTTCATCCCTGTGCAT 4602
792 ysmetCysLeuPheHisIleuThrHisLeuSerAspIleGlyGlyMet 808
4603 CCTCTTGTGCTCGACAGCTCTAGTCTTACATGTTGGGAGACTCCTCC 4652
809 AspTyr..... 810
4653 TTTTACAGAGCTTTTGAACAGACAGAGATATGCTCTTTGAGTAA 4702
811 lLeuValLysSerLeuSerSerGluProCysAspLeuGlnGlnIleGlnL 827
4703 TTTAGAAAGGGCTGTTTCAGCAAAACAGTGCAGTGTAGATGAATATTTT 4752

```

```

827 euValserCysLeuSerAlaAsnAlaVallys.....Ile 839
      |||
4753 GCAGCAGTCTGCAGACCTATGTTGACTCATTAAGGCCGAATTTGGTC 4802
      |||
840 LeuAlaGlnAsnLeuHisAsn.....LeuVallysLeuSerI1 852
      |||
4803 TTGTCAACGATCTGAGAACTTTCAAGGTGACTGTGTAAGAGATGCA 4852
852 eLeuAspLeuSerGlnAsnTyLeu..... 860
      |||
4853 GCTGGGCTTGGAGAGGGCTGTTCCATCCCTGCATCCCTTGTGTGC 4902
861 .....GluLysAspGlnAsnGlnAla 867
      |||
4903 CTCACAGCTCTAGCTTAGACAGTTTGGAGACATCCCTTTTACAGAGCT 4952
868 LeuHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAlaLe 884
      |||
4953 CTTTGTAGAACACAGACGAGATATGCTCTTTGTAGTAATTTAGAGGGGT 5002
884 uMetLeuProTyrGlyCysAspArgIleGlnLysSerLeuSerLeuLeu 901
      |||
5003 TGTTTCAGCAAAACAGTGCAGTGTAGATGATATTTCACAGCATCTGC 5052
901 yshis.....LeuGluGluValProGlnLeuValLysLeuGly 913
      |||
5053 AGCAGCAGATCTGCACGAGAAAGAACCCCTTCGGCCCGCAGGAGGCT 5102
914 LeuLysAsnTyrTrpArgLeuThrAspThrGluIleArgIleLeuGlyAla 930
      |||
5103 GTTGAA.....GAGCTTGACTCCCTCTGTGAGGT 5131
930 ePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyA 947
      |||
5132 GTACCG.....CAGTCCCTCGAGAGAGCTAGAG 5160
947 snArgValSerSerAspGlyTyrTrpLeuAlaPheMetGlyValPheGluAsn 963
      |||
5161 AGAAATGGAAGTCAAGGATTTATGAAAATAGGAATTT..CAGAG 5207
964 LeuLysGlnLeuValPhe.....PheAspPheSerThrly 975
      |||
5208 CTCGACACCTATTATAGTTCTGAAGCAAGAGCTGACTGCCTTAGAGAA 5257
975 sGluPheLeuProAspProAlaLeuValArgLysLeuSerGlnValLeu 992
      |||
5258 GCAGTATTTGTCAGAAAT.....GACACGTGGCAACAGAACCTGA 5298
992 eLysLeuThrPheLeuGlnGluAlaArgLeuValGlyTyrGlnPheAsp 1008
      |||
5299 CAAGCTGACTCTGAGATGAGACTCAAGTTGGCGCAGAAAAAGAACAG 5348
1009 AspAspAspLeuSerVal 1014
      |||
5349 ACGGAACAACCTGCACCTT 5366

```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-16216-2

```

seq_documentation_block:
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA

```

```

; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-2

```

```

alignment_scores:
      Quality: 149.50      Length: 1156
      Ratio: 0.266      Gaps: 52
      Percent Similarity: 48.702      Percent Identity: 18.426

```

alignment_block:

US-09-697-089-2 x PCT-US95-16216-2 ..

Align seg 1/1 to: PCT-US95-16216-2 from: 1 to: 10136

```

39 AsnIleIleCysCysGluLysValGlnGlnAspAlaIleArgIlyIle1 55
      |||
2220 AACGCTTGACAGTAAGTCACTGAGAGACCCAGAA..CTAGC 2266
55 eHisMetIleLeuLysLysGlySerGluSerCys..... 66
      |||
2267 TTATATGAGACTACAGCAAGAGCTGAGTTCTACAGATCAGAAACATCAGA 2316
67 .....AsnLeuPheLysSerLeuLysGluTyrAsnTyrPro 79
      |||
2317 AGCAATATGAAAATATGTTTGAAGACT..... 2345
80 LeuPheGlnAspLeuAsnGlyLys.....SerLeuPheHis.....G1 92
      |||
2346 .....TCTCAGCTTACTGGCAGAGTTGAAGATCTTAGAACACACACTTCA 2389
92 nThrSerGluGlyAspLeuAspAspLeuAlaGlnAspLeuLysAspLeuT 109
      |||
2390 GTTACTGTCAAAATGAATATATGACCAAGACCGGTATACCAAGACTTGC 2439
109 yHisThrProSerPheLeuAsnPheTyrProLeuGlyGluAspLeuLeas 125
      |||
2440 ATGCCGAATATAGAGACCTCAGGAGATCTGCTAAATCCAAAGATCTCT 2489
126 IleIlePheAsn.....LeuLysSerThrPheThrG1 136
      |||
2490 CTGGGACAAATGAGAGATCATCAGAGAGTCTTTGGCTTTGATCAGCA 2539
136 uProValLeuTyrPargLysAspGlnHisHisArgValGluGlnLeuT 153
      |||
2540 GCCTGCATG.....CATCATCTCTTGCAAATATATATG 2574

```



```

4153 GATTCACAGCACTAGAGAAATGGCAGAGAGGTAGGAAACTACTAAT 4202
714 ILeTyrSerLeuMetValGluAlaSerProLeu.....Thr1 726
4203 GAAGTTAAATATTAATGATGACAGTGGCTTCTCCATGTCAGTTAG 4252
726 egluAsp.....GluArgHis1 732
4253 GGAAGACATACACAGAGGTGAATTTGGTGACACCAACCAATGACACACC 4302
732 LeThiSerValThiAsnLeu.....LysThiLeuSer 742
4303 CTGTGCTTTGGCTCCATTGGACAGAGTAATCTTCAAGCAGCCTTGACA 4352
743 ILeHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThiAspSer1e 759
4353 TTGTTCAGACAAAGAGACTCAATGCACTTTGGCCGAATTCAGAGAAAT 4402
759 ugiLysAsnLeuLysAsnLeuThiLysLeuLeuMetAsp..... 771
4403 CTATCTTTACAAAGTGAACACAAATTTTACATGATCAGCACTGTCCAGA 4452
771 ..... 771
4453 TGAGCTCTAAATGTCAGAGCTGCAGACCTATGTCATCTAATAAGGCC 4502
772 .....AsnIleLysMetAsnGluGluAspAla11 781
4503 GAAATTTGGCTGTTCACAGAACTTGAGAAACTTTCAAGGTGACTTGGCT 4552
781 eLys.....LeuAlaGluGlyLeu...LysAsnLeuLysL 792
4553 GAAGGAGATGCAGCTGGCTTGAGAGAGGGCTCGTTCCATCCCTCTCAT 4602
792 ySMeLcysLeuPheHisLeuThiHisLeuSerAspIleGlyGlyMet 808
4603 CCTCTTGTGCTGCACAGCTCTAGCTTAGCAGTTGGAGACTCTCTCC 4652
809 AspTyr..... 810
4653 TTTTACAGAGCTTTTACACAGACAGAGATATCTCTTTTGAGTAA 4702
811 ILeValLysSerLeuSerSerGluProCysAspLeuGluGluLeu1 827
4703 TTTAGAAAGGGCTGTTTCAGCAACAGAGTCAGTGAATGAATTTT 4752
827 euValSerCysLeuSerAlaAsnAlaValLys.....Ile 839
4753 GCAGCAGCTCTGCAGACCTATGTTGACTCAATTAAAGCCGAAATTTGCTC 4802
840 LeuAlaGlnAsnLeuHisAsn.....LeuValLysLeuSer11 852
4803 TTGTCAACAGAACTTGAGAACTTTCAAGGTGACTTGTGAAGAGATGCA 4852
852 euLeuAspLeuSerGluAsnTyrLeu..... 860
4853 GCTGGGCTTGAGAGAGGGCTGTTCCATCCCTGTCACTCTTGTGTC 4902
861 .....GluLysAspGlyAsnGluAla 867
4903 CTCACAGCTCTAGTCTTACAGAGTTGGAGACTCTCTTTTACAGAGCT 4952
868 LeuHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThiAla1e 884
4953 CTTTGAACACAGACAGATATGCTCTTTTGAGAAATTTTGAAGGGGT 5002
884 uMetLeuProTrrGlyCysAspValGlnGlySerLeuSerSerLeuLeu1 901
5003 TGTTCAGCAACACAGTGCAGTGAATGAAGATATTTTGCACAGACTC 5052
901 yHis.....LeuGluGluValProGlnLeuValLysLeuGly 913
5053 AGGAGAGAAATCTGACACAGAAACCCCTTGGCCCCCAGCAAGGGT 5102

```

```

914 LeuLysAsnTrrArgLeuThiAspThiGluIleArgIleLeuGlyAla1ph 930
5103 GTTGA.....GAGCTTGAAGTCCCTCTGTGAGCT 5131
930 ePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyA 947
5132 GTACCG.....CAGTCCCTCGAGAGCTAGAG 5160
947 snArgValSerSerAspGlyTrrPheValAlaPheMetGlyValPheGlu 963
5161 AGAAATGCAAGAGTCAAGGATATGAAAAATAGGAATT...CAAGAG 5207
964 LeuLysGlnLeuValPhe.....PheAspPheSerThiLys 975
5208 CTCGACAGCTTATTAAGTTCTGAAGCAGACAGCTTGACTTGTAGCA 5257
975 sgluPheLeuProAspProAlaLeuValArgLysLeuSerGlnValLeu 992
5258 GCAGTATTTCAGAAAT.....GACAGTGGCAGACAGAGCTGA 5298
992 eLysLeuThiPheLeuGlnGluAlaArgLeuValGlyTrrPheAsp 1008
5299 CAAGGCTGACTCTGGAGATGAGTCCAAAGTTGGCGCAGAAAGAACAG 5348
1009 AspAspAspLeuSerVal 1014
5349 ACGGAACAACCTCTCACTT 5366

seq_name:/cgn2_6/plodata/2/lna/6A_COMB.seq:us-08-947-823-1
seq_documentation block:
; Sequence 1, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Isqouhi
; APPLICANT: Yashoobi, Jafar
; APPLICANT: Bodeau, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Conferring
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-Oct-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-Oct-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-Oct-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:

```



```

19376 ATTTGGTCCCGAATTGGATTGCTTACTGACTAGAACTCTGTGTGA 194255
19426 GGT...TTTAAAGTCTCAAC.....ACAACCA 19451
944 uAlagIyAsnArqVAlSerSerAspGlyTrpIeuAlaPhetGlyAlp 961
19452 CTGTGGGCTCTGTGTGCACAAATCGCCGCGGATTTT...CACTTCC 19498
961 heGlyAsnIeuGlySInIeuValaPhetAspPheSer...ThIrySGlu 976
19499 CTTCAAATTTGAAAGACGTGTTGTGTAGACTTTCCTGTGACATCCGAT 19548
977 PheIeuProAspProIaIeuValArqVAlSleSerGlnValIeuSerly 993
19549 TCACTrTTCACACATAGACGAGACTGCCCAACCTTGAAATTTTCCCTTTA 19588
993 sleuTrPheIeuGlnGluAlaArqVAlGlyTrpGlnPheAspAspa 1010
19599 TGATACATCATTCACAGGAGAGAA.....TGAAACATGGGGAGG 19639
1010 spAsp 1011
19640 AAGAC 19644

seq_name: /cgn2_6/ptodate/2/ina/6A_COMB.seq:US-08-947-823-2

seq_documentation_block:
Sequence 2, Application US/08947823
Patent No. 6114605
GENERAL INFORMATION:
APPLICANT: Williamson, Valerie M.
APPLICANT: Kaloshian, Isgouhi
APPLICANT: Yaghmoobi, Jafar
APPLICANT: Bodeau, John
APPLICANT: Milligan, Stephen
TITLE OF INVENTION: Procedures and Materials for Confering
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/947,823
FILING DATE: 09-OCT-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18802
FILING DATE: 09-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastlian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-070210US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

```



```

415 rValaAsnGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyrThrA 432
      ::::: |||||
1980 TACTACTACATGGATGGATTAACA..... 2004
432 LacInaRphLeuPheLysProLysPhePheHisLysSerPheGlnGlu 448
      :::: ||| |||
2005 .....AGACCTTTTCTGTANA 2019
449 TyrThrAlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProG1 465
      ||| ||||| ::::: ||| |||
2020 TCTAAGAAAGGAAGTAGAGATT.....ATTGTGACAACTCGGGAAGAA 2063
465 uGluValThrLysGlyAsnGlyTyrLeuGlnLysMet...ValSerHis 481
      | ::::: ||||| :::::
2064 AGTGGCTTTGCATGGAAAGCTGACACGTACCTCTTGACCTTCGATTGC 2113
481 erAspIleThrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySer 497
      ::::: ||||| ::::: |||
2114 TAAGACCAATGAAAGTTGGAACTATTAGAGAAAGGCAATTGGGAT 2163
498 Ser.....ValGluAlaThrArgAlaValMetLysHis 508
      ::::: ||||| ::::: |||
2164 GAGAGTTGCCCTGATGAACTATTAGATCGGTTAAAGAAATAGCGAANA 2213
508 s.....LeuAlaAlaValTyrGlnHisGlyCysLeuLeuG 520
      | ::::: ||||| ::::: |||
2214 TTGTAAAGGCTTCCTTTGGTGGCTGATGATGCTGGAGTCATTTGCTG 2263
520 LysLeuSerIleAlaLysArgProLeuThrParg.....GlnGluSerLeu 534
      || :::: ||||| ::::: |||||
2264 GGAGGGAA...AAGAAAGAGAGTGTGTGGCTGAGTTCAAGTACTTTGG 2310
535 GlnSerValLysAsnThrThrGlnGlnGluLeuLysAlaIleAsnI1 551
      ::::: ||||| ::::: |||||
2311 ACTTCTTTATTTTGAAAGTAGAGTGAAGTGAAGTGAAGTATAGAAAT 2360
551 eaAsnSerPheValGlyCysGlyIleHisLeuTyrGlnGluSerThrSerL 568
      :::::
2361 AAGT..... 2364
568 ySerAlaLeuSerGlnGluPheGlnAlaPhePheGlnGlyLysSerLeu 584
2364 ..... 2364
585 TyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPheGlnI1 601
      ::::: |||
2365 .....TATGACCA 2372
601 sLeuProAsn.....CysAla..... 606
2373 TTTACCAATCACTCAAGCCATGCTGTGATTTTGCAGATTTTCCGA 2422
607 .....SerAlaLeuAspPheIleLysLeuAspPheTyrGlyLysAlaMet 621
      ::::: ||||| ::::: |||||
2423 AGGACACTTCATTTGACAACTATGATGATTTATTTGGTGTGGA 2472
622 AlaSerTyrGluLysAlaIleGlnAspThrGlyGlyIleHisMetGluG1 638
      ::::: ||||| ::::: |||||
2473 GGATTTTGGGAAAGACGAGATGAACT.....ATGGAANA 2510
638 uAlaProGluThrTyrIleProSerArgAlaValSer.....LeuP 652
      | ::::: ||||| ::::: |||
2511 AGTGGGAAGATTTATATGATGATTTAATTACAGTACCTGGTAATTT 2560
652 hepPheAsn.....TyrLysGlnGluPheArgThrLeuGluValThr 665
      ||||| ::::: ||||| ::::: |||
2561 GTTTCAATGAGATAGGTATGACATGAAATTTCCAAATTCATGATCTT... 2607
666 LeuAlaArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyL 682
      ::::: ||||| ::::: |||
2608 GTGCATGACTTTTGTGATTAAGCAAGAAAGAAATTTGTTGATCA 2657

```

```

682 sIlePheSerSerAlaThrSer.....LeuArgLeuGlnIle..... 694
      ::||| ||||| ||| ||| |||||
2658 GATTAAGATCAAGTGCCTCCATGATTTGTGCTGTCAAAATTTACCATTG 2707
694 ..... 694
2708 ATTGTGATGAGGAGGACACTTTGGCTTAATTTGTGATGTTGCATTCA 2757
695 .....LysArgCysAlaGlyAlaIleGlySerLeuSerLeuValLeuSe 709
      ||||| ::::: |||
2758 AATAAGAAAGGCAATCTGCT..... 2778
709 rThrCysLysAsnIleTyrSerLeuMetValGluAlaSerProLeu... 724
      ||||| ::::: ||||| ::::: |||
2779 .....AAACCTCTATTCTTTGAGCATATTGGAGACCACTGGATG 2821
725 ..ThrIleGluAspGluArgHisIleThrSerValThrAsnLeuLysThr 740
      ::::: ||||| ::::: |||||
2822 ACAGTGTTCGTGATGCATTTGCCTAAGACACTTGAGGCTCTTGTAGAGT 2871
741 LeuSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAs 757
      ||||| ::::: |||
2872 TTGGACCTGCAT..... 2883
757 pSerLeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleL 774
      ::::: |||||
2884 .....ACGCTTTTATTCATG..... 2898
774 yMetAsnGlnGluAspAlaIleLysLeuAlaGluGlyLeuLysAsnLeu 790
      ::||| ::::: |||
2899 .....GTGAAGATTCCTTG.....CTG 2916
791 LysLysMetCysLeuPhe...HisLeuThrHisLeuSerAspIleGlyG1 806
      ::::: ||||| ::::: |||||
2917 AATGAATAATGCATGTGATCATCTGTAAGTGAAGTGAAGTGAAGTGAAGT 2955
806 uGlyMetAspTyrIleValLysSerLeuSerSerGluProCysAspLeuG 823
      ::||| ||||| |||
2956 ....ATTGACACACAAGTTAAATATCTG.....CCTTTGCTTTCT 2992
823 LysGluIle...GlnLeuValSerCysCysLeuSerAlaAsn...AlaVal 837
      ::::: ||||| ||| ::::: |||||
2993 CAACCTCTGGAATGTAGAAAGCTGTTGTGTACCAACACAGATCATTC 3042
838 LysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIleLeuAs 854
      ::||| ::::: ||||| ::::: |||||
3043 TTGGTACTATTACCGGAATTTTGGATCTGTGAAGTGGCGAGTCTGTCTC 3092
854 pLeuSerGlu...AsnTyrLeuGluLysAspGlyAsnGlnAla..... 867
      ::::: ||||| ::::: |||||
3093 CGTGGATGCTGTCTTTCTTTGATGATGATGACATGAAATCAATATTTGA 3142
868 .....LeuHisGluLeuIle 872
3143 TAGCAGAGGACACAAAGTTAGAGACTTGAGAAATTTAAGCGAACTGTG 3192
873 .....AspArgMetAsnValLeuGlnGluLeuThrAlaLe 884
      ::||| ||| ::::: |||||
3193 ATTTCCTATTGGAAGATCAAAAGAAATTTTCAAAAGTTTCCCAATCT 3242
884 uMetLeu.....ProTyrGlyAspValGln 894
      | ::||| |||
3243 TCAGTTGCTTCATTTGAACTCAAGGATCATGGATTTATTTCAACAGAGC 3292
894 LysLeuSerSerLeuLeuLysHisLeuGlnGluValProGlnLeuVal 910
      ||| ||| ||||| ::::: |||
3293 AACATGTGTTCTCGGAATTTGGATTTTCCTAACGTGAACTA...GAAACACTC 3339
911 LysLeuGlyLeuLysAsnTyrArgLeuThrAsp..... 921
      ::||| ::||| ::|||
3340 TCTGTAGGTTTAAAGTTCAACACAAACGATAGTGGTCTCTGTAGAC 3389
922 .....ThrGluIleArgIleLeuG 928

```

```

3390 GACAAATCGCCGCTGGATTTCCTCCCTCAAAATTTGAAATATCTG. 3438
928 lylalaphhegilylysasnPro..... 925
3439 .....TGGTGGCTGAATTTCCCGCTGCATCCGATTCACATATCAACATA 3483
936 .....LeuylsasnpheglinglnleuAsnleuAlaGlyAsnArgylSe 950
3484 GCGAGACTGCGCAACCTTGAAAGATTGTCCCTTTATCATCAATCAATCA 3533
950 rSerAspGlyTTP...leuAlaphemetGlyValPhegluAsnleuLysG 966
3534 TGGAGAGATGGAACATGGGAGAGAACACCTTTGAGAATCTCAAAAT 3583
966 lnleuValPhepAspPheSerThrlysgluPhe 977
3584 TTTTGAACCTTCATCAAGTTAGTATTTCCAAGTGG 3618

```

seq_name: /cgn2_6/ptodata/2/lna/6B.COMB.seq:US-09-180-439-5

documentation_block:

Sequence 5, Application US/09180439

```

GENERAL INFORMATION:
PATIENT NO. 6225532
APPLICANT: Dixon, Mark S
APPLICANT: Halzixanthlis, Kostas
APPLICANT: Jones, David A
APPLICANT: Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
FILE REFERENCE: 620 - 53
CURRENT APPLICATION NUMBER: US/09/180,439
CURRENT FILING DATE: 1998-12-06
EARLIER APPLICATION NUMBER: PCT/GB97/01249
EARLIER FILING DATE: 1997-05-08
EARLIER APPLICATION NUMBER: GB 9609681.3
EARLIER FILING DATE: 1996-05-09
EARLIER APPLICATION NUMBER: GB 9619924.5
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 3541
TYPE: DNA
ORGANISM: Lycopersicon esculentum
US-09-180-439-5

```

```

Alignment_scores:
Quality: 143.50      Length: 379
Ratio: 0.755         Gaps: 17
Percent Similarity: 50.132 Percent Identity: 23.747

```

alignment_block:

US-09-697-089-2 x US-09-180-439-5 ..

Align seg 1/1 to: US-09-180-439-5 from: 1 to: 3541

```

675 GluAspIleThrlyLeuGlyLysIlePheSerSerAlaThrSerLeuAr 691
1227 GAAGAAATAGCTTACTTA.....AGGTCTCTTACTTAAGCTATC 1264
691 GluGlnIleLysArgCysAlaGlyValAlaGlySerLeuSerLeuVal 708
1265 TTGGATATCAAC.....TTTCTTAGTGTTCATTCCTGCTTCAT 1305
708 euserThrCysLysAsnIleTyrSerleuMetValGluAlaSerProLeu 724
1306 TGGGGAATTCGAACAACTGCTTTTGTATCTTTCAATAATCAAGCTT 1355
725 .....ThrIleGluAspGluArgHisIleThrSerValThrAsnLeu 739
1356 TCTGGCTCTATCTCTGAAGA.....ATAGCTTACTTAAG 1390

```

```

739 sThrLeuSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeu 756
1391 GTCTCTTACTTACCTAGATTGTAAGAGAAATGCTCTTAATGGCTATTC 1440
756 hArgSerLeuGlyAsnLeuLysAsnLeuThrlyLysLeuIleMetAsp 772
1441 CTGCTTCATTTGGGGAATCTGAACAACTGTCTAGGTGTGATCTTACAAT 1490
773 IleLysMetAsnGluGlnAspAlaIleLysLeuAlaGlyLeuLysAs 789
1491 AATCAGCTTCTTGGC.....TCTATTCTGGAAGAAATAGCTTA 1528
789 nLeuLysLysMetCysLeuPheHisLeuThrHisLeuSerAspIleGly 805
1529 CTGAGATCTCTTACTAATCTATATTGGGTATTAATCTCTTATTGGAC 1578
806 .....Glu 806

```

```

1579 TTAATCTGCTTCATTCGCAATATGAGAAATCTGCAAGCTGTTTCTC 1628
807 GlyMetAspTyrIleValLysSerLeuSerSerGluProCysAspLeuG 823
1629 AATGATTAACATCTCATTTGGGAAATTCCTTCAATTGTGTGCAATTAAC 1678
823 uGluIleGlnLeu.....ValS 829
1679 ATCATAAGAACTGTGTATATGCCGAGAAACAATTTGAAGGCAAGTTC 1728
829 erCysCysLeu.....SerAla 834
1729 CGCAATGTTGGGTAATATACAGTACCTCTGCTTTGCAATGTCATCT 1778
835 AsnAlaValLys...IleLeuAlaGlnAsnLeuHisAsnLeuValLysLe 850
1779 AATAGTTTCAGTGGAGAGCTCCCTTCATCTATTTCCAAATTTAAATCAG 1828
850 uSerIleLeuAspLeuSerGluAsnTyrLeuGluLysAsp..... 863
1829 AAAAATACTGATTTGGCGAACAATCTCGAGGAGCAATACCACAAT 1878
864 ....GlyAsnGluAlaLeuHisGluLeuLeuAspArgMetAsnValLeu 878
1879 GTTTTGGCAATATATATAGCTCCAGCTTTTGTATATGAGCAAT..... 1922
879 GluGlnLeuThrAlaLeuMetLeuProTyrGlyCysAspValGlnLys 895
1923 AACAACTTCTGGG...ACTCTCCAAACAATTTTATGATTTGATGTTTC 1969
895 rLeuSerSerLeuLysHis..... 902
1970 ACTGATTAAGTCTCAACTTGCATGCATGAATGAATGAAATCCCTT 2019
903 ....LeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsnTyr 917
2020 GTCTTTTGACAATTCGCAAAAGCTCAAGTCTGTATTTAGGAGACAAT 2069
918 ArgLeuThrAspThr.....GluIleArg 925
2070 CAACCTCAACACACATTTCCCATGCTGTTGGAACTTTGCCAGAGCTGAG 2119
925 gIleLeuGlyAlaPhePheGlyLys.....AsnProLeuLys..... 937
2120 AGTTTAAAGTTGACATCGAATAAATTGCATGAGCACTATATAGATCATCAG 2169
938 .....AsnPheGlnGlnLeuAsnLeuAlaGlyAsn 947
2170 GGGCTGAATCATGTTCCATGATCTTGATCATATCATATCTCTCCCAAT 2219
948 ArgValSerSerAspGlyTyrPheAlaPheMetGlyValPheGluAsnLe 964
2220 GCATTCTCGCAAGAC.....TTACCAACAAGCTATTTTGAACATTT 2260

```



```

1830 GATCAAGATATGACGACAAAGATTGTTGATACAAATTTTCAGTCAAGT 1879
536 erval..... 537
1880 TAGTGGCTCATTCAAATTTGAGTGAATATTGATGTTGCGTAAAT 1929
538 .....LysAsnThrThrGluGluIleLeuLysAlaIleAsnIleAs 552
1930 TGGGAAACAAACGTGTTGAAAGAGTATCTTAATG..... 1965
552 nSerPheValGluCysGlyIleHisLeuTyr..GlnGluSerThr.... 566
1966 .TCCTTAGATGATGTGTGGTACTACTACATTTGGATGAGTTGACAAAGACC 2014
567 .....SerLysSerAlaLeu...SerGlnGluP 575
2015 TTTTCCGAGCTAAGAAAGAACTAGGATTTATTTGACAACTGACAGAA 2064
575 heGluAlaPhePheGlnGlyLysSerLeuTyrIleAsnSer..... 588
2065 AGGAAGTGGCTTTGATGAGAAAG.....CTGACACTGATCTCTT 2105
588 ..... 588
2106 GACCTTCGATTGCTAAGACAGATGAAAGTTGGACCTTTAGATAAAG 2155
589 .....GlyAsn.....IleProAspTyrLeuPheAspPheGluH 601
2156 GACATTTGGTATGATGAGAGTGGCCCTGATGAACTTTAGATGTGCGTAAAG 2205
601 ILeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
2206 AAATAGCCGAAATTTGAAAGGCTCTCTTGTG.....GCTGATCTGATT 2252
618 GlyCylAlaMetAlaSerTTPGluLysAlaIleGluAspThrGlyIle 634
2253 GCTGGACTCATCTGCTGGAGGAGAAAG.....AAAGAGTGTGTGCT 2296
634 eHsMeGluGluAlaProGluThrTyrIleProSerArgAlaValSerL 651
2297 TGAAGTCAAGTAGTTGAGTCTTTATTTTGAACAGTCAAGTGGAGAG 2346
651 eu..... 651
2347 TGATGAAAGTTATGAAATTAAGTATGACCATTTACACATCACTCAAG 2396
652 .....PhePheAsnTrp..... 655
2397 CCATGCTTCTTCACTTTCGCAAGTTGGCCGAGAGACACTCCTTTGACAT 2446
656 ..... 656
2447 CTATTTGTTACTGTTATTTGGGCTGGAAGATTTGTGAAAGAGACGG 2496
658 IuPheArgThrLeuGluValThrLeuArgAspPheSerLysLeuAsnLys 674
2497 AGATGAGAGGCTATGAGAGAGTGTGAG..... 2525
675 GlnAspIleThrTyrLeuGlyLysIlePheSerSerAlaThrSerLeuAr 691
2526 .....ATTATATGATGATTTAATTTCCAGTAGCTTGTAATT.. 2564
691 gLeuGlnIleLysArgCysAlaGlyValAlaGlySerLeu..... 704
2565 .....TGTTCATGAGATAGTGATATACTGAAATTTCC 2598
705 .....SerLeuValLeuSerThrCys.....Lys 712
2599 AAATTCATGATCTTGTGCATGACCTTTGTTGATAAAGCAAGAAAGGA 2648
713 AsnIleTyrSerLeuMetValGluAlaSerPro..... 723
2649 AATTTGTTGATCGATAGATCAAGTGCCTCATCAGATTTGTTGCCCTCG 2698

```

```

724 ....LeuThrIle.....GluAspGluArgHis..... 731
2699 TCAAATATCCATTCATATATGATGAGGAGGAGACACTTTGGGCTTAAT 2748
732 ..IleThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGln 747
2749 TTGTCACTGTTGCAATTCAAATAGCAAAAGGCATCTGTGTAACACTTAT 2798
748 AsnGlnArgLeuProGlyGly..LeuThrAspSerLeuGlnLysLeu 763
2799 TCCTTGGAGATAAATGAGACACAGCTGATGACAGTCTTCTGATGCATT 2848
763 sAsnIleThrLysLeuIleMetAspAsnIleLysMetAsnGluLysAs 780
2849 TCACCTAAGACACTTGGAGCTTATAGAGTGTGAGACTGGAACCCCTT 2898
780 IalIleLysLeuAlaGluGlyLeuLysAsnLeuLysLysMetCysLeuPhe 796
2899 TAATCATGCTGAATGATTTCTTTG.....CTGAATGAAATATGATGCTGTG 2942
797 ..HisLeuThrHisLeuSerAspIleGlyGluGlyMetAspTyrIleVa 812
2943 AATCATTTGAGAGTACTTACAG...ATTGCGACACACATTAATATCTGCC 2969
812 LysSerLeuSerSerGluProCysAspLeuGluGluIleGlnLeuValS 829
2990 TTTCTCTTCTTCAAC...CTCTGCAATCTAGAAAGTCTG...TTTGTGT 3033
829 erCysCysLeuSerAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHis 845
3034 CT.....AACAAAGATCAATCTTGGTACTACTACGAGAAATTTTG 3074
846 AsnLeuValLysLeuSerIleLeuAspLeuSerGlu...AsnTyrLeuG 861
3075 GATCTGTAAAGTTGCGAGTGTCTCCGTGGTCTTCTTCTTTCTTGA 3124
861 uLysAspGlyAsnGluAlaLeuHisGluLeuIleAspArgMetAsnVal 878
3125 TATGATGCGAGATCAATCAATA.....TTGATATGCAAAAGACCAAAAGT 3168
878 euGluGlnLeuThr.....AlaLeuMetLeuProTTPGlyCysAsp 891
3169 TAGAGAACTTGAGAAATTTAGGGGCACTGTGATTTCTATTCGAAAGAT 3218
892 ValGlnGlySerLeuSerSerLeuLeuLysHisLeuGlnGluValProG 908
3219 ACAATG.....AATATTTCAAAAGGTTTCCCAATCTT...CA 3253
908 nLeuValLysLeuGlyLeuLys..AsnTyrParg..... 918
3254 GGTGCTTCAGTTGACACTCAAGGAGTCACTGGATTTATCAACAGAGCAAC 3303
919 .....LeuThrAspThrGluIleArgIleLeu 927
3304 ATTGCTTCCCGAAATGATGATTCCTTAAGTCACTGAAACACCTCTGTGA 3353
928 GlyAlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLe 944
3354 GGT.....TTTAAAGTTCAAACACCAACCAACCTG 3382
944 uAlaGlyAsnArgValSerSerAspGlyTTPLeuAlaPheMetGlyValP 961
3383 TGGCTCTCTGTTGTGACAAATGCGCGTGGGATTTTCACTTCCCT.... 3428
961 heGluAsnLeuLysGlnLeuValPhePheAspPheSer...ThrLysGlu 976
3429 ..TCAAATTTGAAAGACCTGTTGTTATGACTTCTCTGACATCCGAT 3476
977 PheLeuProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLy 993
3477 TCACATATCAACATATAGCGAGACTGCCCAACCTTGAAATTTGTTGCCCTT 3526

```



```

1122 GAAACTT.....CACCTGGGTTTGCTGT 1147
756 hrAsperleuglyasnLeuLyAsnleuthrLysleuIImetAspSn 772
      ::::|||||:|||||:|||||
1148 TCGAGAAATCATTAGGAACCTTCACAACACTTAT.....TTAAGGCAT 1168
773 lLeuMetAsnJucLuAsPaLaileLysLeuAlagLuclyLeuLysAs 789
      :::: :|||: :|||: :|||
1189 AATGACATAGAGGCTTCTGACTGCTGAGTCTGCA.....CTCAAAA 1232
789 nLeuLysLysMetCysLeuPheHisleuthrHisLeuSerAspIlegLg 806
      ||||: :||: :|||: :|||: :|||
1233 CCGTCCCACTTCGAACACTTAACCTGACGCCAACATGAGCGCTTGCT 1281
806 luclyMetSprylrLeValLysSerLeuSerGluProCys.. Asp 821
1282 .....CTCCAGAATGAGCATTCCAAAGAATGCTCCAG 1314
822 LeuGlucInlIleGlnLeuValSerCysCysLeuSerAlasnaIvalLy 838
      ||||| :|||: :|||: :|||
1315 CTAGAACTCTCGATTGGCATTTACC CGCTTACACACTTAATGCTCCACA 1364
838 sIleLeuAlagInAsnLeuHisAsnLeuValLysLeuSerlle..... 852
      : |||||: ||| :|||: :|||
1365 AAGTCCCTTCGAACACTTCATCTTCCTCAAGTTCTGATCTCACTTACT 1414
853 ....LeuAspLeuSerGlu..... 857
1415 GCCTTCCTGTATCACCAACATCAGCATCTTACAGAGCTACACAGTTCTC 1464
858 AsnTyrlEuGluLyuAsprglYAsnGluAlaLeuHisGluLeuIleAspAr 874
1465 CGGCATCTCAACTTAAAAAGGAATCACTTTCAAGATGGAGCTATCCAGAA 1514
874 gMeAsnValleuGluGlnleuthralaleu.....MetLeuProTrpg 889
      : ||||: :|||: :|||: :|||
1515 GAACAACCTACTTCAAGACCGTGGGACAGCTTGAGAGTTCTGATTTGGTCT 1564
889 lYcAspApyAlnglySerLeuSerSerLeuLeuLysHISleuGluLu 905
      :||| :||| :||| :|||
1565 CTGTGT.....GGTCTCTCTATTAAGCAGCAGCAAGCATTTCCAGAGC 1605
906 ValProGlnleuValLysleuGlyLeuLysAsnDTPargyleuthrrAspTh 922
      :||: :|||: :|||: :|||
1606 TTGGGAAAAATGAGCATGTGACTTAAAGCACACAGCTGACATCGGA 1655
922 rgUlnIeArgyleLeuGlyAlaPhePheGlyLysAsnProLeuLysAnp 939
      :|||: |||: :||: :||
1656 CAGCATTTGATTCCTTAGCCATCTTAAGGA..... 1686
939 heGlnIleuAsnLeuAlaGlyAsnArgValSer 950
      |||||: |||: :||| :|||
1687 ..ATCTACCTCAATCTGGCTGCTCAACAGCATTTAAC 1719

seq_name: /cgn2_6/p/tdata/2/ina/5B_COMB.seq:US-08-833-823-3

seq_documentation_block:
# Sequence 3, Application US/08833823
# Patent No. 5969093
# GENERAL INFORMATION:
# APPLICANT: Jacobs, Kenneth
# APPLICANT: McCoy, John
# APPLICANT: Kelleher, Kerry
# APPLICANT: Carlin, McKough
# TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
# TITLE OF INVENTION: ENCODED THEREBY
# NUMBER OF SEQUENCES: 12
# CORRESPONDENCE ADDRESS:
# ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs
# STREET: 87 Cambridgepark Drive
# CITY: Cambridge
# STATE: Massachusetts
# COUNTRY: USA
```

```

1      ZIP: 02110
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Floppy disk
4      COMPUTER: IBM PC compatible
5      OPERATING SYSTEM: PC-DOS/MS-DOS
6      SOFTWARE: PatentIn Release #1.0,
7      CURRENT APPLICATION DATA:
8      APPLICATION NUMBER: US/08/833,823
9      FILING DATE: 10-APR-1997
10     CLASSIFICATION: 530
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: 08/514,014
13     FILING DATE: 11-ANG-1995
14     ATTORNEY/AGENT INFORMATION:
15     NAME: Brown, Scott A.
16     REGISTRATION NUMBER: 32,724
17     REFERENCE/DOCKET NUMBER: G16000
18     TELECOMMUNICATION INFORMATION:
19     TELEPHONE: (617) 498-8224
20     TELEFAX: (617) 876-5851
21     INFORMATION FOR SEQ. ID NO. 3:
22     SEQUENCE CHARACTERISTICS:
23     LENGTH: 2582 base pairs
24     TYPE: nucleic acid
25     STRANDEDNESS: double
26     TOPOLOGY: linear
27     MOLECULE TYPE: cDNA
28     HYPOTHETICAL: NO
29     FEATURE:
30     NAME/KEY: CDS
31     LOCATION: 52..2034
32     OS-08-833-823-3

```

alignment_scores:	142.00	length:	545
Quality:	0.553	Gaps:	24
Ratio:	47.156	Percent Identity:	22.936
alignment_block:			
US-09-697-089-2	x	US-08-833-823-3	..

Align seq 1/1 to: US-08-833-823-3 from: 1 to: 2562

486 ThrTyrSerSerLeuLeuArgTYrThrCysGlySerSerValGluAlaIarh 502D
:::||||| ::||| ::||| ::||| ::|||
265 ACCCTTCACACATCATCATCAATAACTTACC.....TTTGTGGATTAAAC 305S
502 rargAlaValMetLys.....HisLeuAlaIaValTyGlnHisGlyc 517H
|||||::: ::||| ::||| ::|||
306 TAGGTGCCAGATTAACTGGATACATGAAGACACTTTCAAAGCCATATATC 355S
517 yslLeuGLyLeuSerlleAlaIalysarProleu.....TPArGln 531H
||| ||| ::||| ::||| ::|||
356 AATTAGACACACTGTGTTAAGTACGGAAATCCCAGATATTCATGCAGAA 405O
532 GluSerleuGlnSerValLysAsnThrThrcLglnGlnIuleuLysAl 548R
|||||::: ::||| ::|||
406 ACATCGTTAATGGGCCCAATCA.....CTGAAG.. 435S
548 aileasnIleasSerPheValGluCysGlyIleHisLeuTYrGlnLus 565S
436CATCTTTTCTTAATCC 451S
565 eRThtSerLySerAlaLeuSerGlnGluPhe.....GIuaIaPhe 578R
::::|::| ||| ||| ||||| ::||| ::|||
452 AAACCGGAATTCCTCAATCC.....GAGTTTATCCAGTGCAACAACCTG 495S
579 PheGlnGlyLysSerLeuTYrIleAsnSerGlnValIleProAspTYrLe 595O
::: ::||| ::||| ::||| ::||| ::|||
496 GAAAACCTTGGAAAGCTTATCTTGGAAAGCACCATATTTCTTCATTTAA 545S
595 upheaSphePheGlnHisIleuProAsnCysAlaSerAlaLeuAsphei 612D

546 GTTC.....CCCAAGACTCCAGCAGCAATCTGA 577
612 IelysleuaspherlyrglyAla..... 620
578 AAGTACTGGATTTTCAGAAATGCTATACACTCTCTAGAGAAAGC 627
621 MetalaSerTrpGluLysAla.....AlaG1 629
628 ATGAGGTCTCTGGAGCAGGCCATCAACCTAAGCCTCAACTCAATGCGAA 677
629 uAspThrIglyIleHisMetGluLysAlaProGluThrTyrIleProS 646
678 TAAATGTAAAGSTATTGAGCTTGGGCTTTTATTCACAGGCTTCCAAA 727
646 er.....ArgAlaValSerLeuPheAsnTrpLys 656
728 GTTGAACCTTTGGAGAACCTCCAAATTGCTGTATATATCAATGCTCG 777
657 GluLupheArgThrLeuGluValThrLeuArgAspPheSerLysLeuS 673
778 CAGAACTCTACTACAGTCTCTGCGCTGGGAGACATTGGAGACATTGA 827
673 nLysGlnAspIleThr.....TyrLeuGlyLysIlePheSerSera 687
828 TGACGAGATATTAGTACGCCATGCTCAAGGACTCTGAAATGCTCG 877
687 IaThrSerLeuArgLeuGlnIleLysArgGlyAlaLysIleSer 703
878 TTGAGAGCTCAACCTGCAGAACACCGCTTCTGACATCTCATCCACC 927
704 Leu..... 704
928 ACATTTCAGTCTTCACCCACTCCAGAAATTGATCTGACAGCAACTCA 977
705SerL 706
978 CTTGAAGGGTTACCCCTGTGGATGAGGGTCTGAACCTGCTCAAGAAAT 1027
706 euValLeuSerThrCysLysAsnIleTyrSerLeuMetValGluAlaSer 722
1028 TAGTCTCAGTGA.....AATCATTTTCGATCAATTTGTCAAAATCACT 1071
723 ProLeuThrIleGluAspGluArgHisIleThrSerValThrAsnLeuLys 739
1072 GCTGCCAATTTCCCTCCCTTACACACTCTACATCAGAGGCAACGTGA 1121
739 sThrLeuSerIleHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuT 756
1122 GAACCT.....CACCTTGGCTTGGCTGCT 1147
756 hrAspSerLeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsn 772
1148 TGGAGAAACTAGGAAACCTTCAGACACTTGAT.....TTAAGCCAT 1188
773 IleLysMetAsnGluLysAspAlaIleLysLeuAlaGluGlyLeuLysAs 789
1189 AATGACATAGAGGCTTCTAGCTGCTGCA.....CTCAAAA 1232
789 nLeuLysLysMetCysLeuPheHisLeuThrHisLeuSerAspIleGlyG 806
1233 CCGTGCCCACTTGCAAACTTAAACCTGAGCCCAATGAGCCTCTGCT 1281
806 lUglyMetAspTyrIleValLysSerLeuSerGluProCys...Asp 821
1282CTCCAGAGTCAAGGCATTCAGAAAGATGCTCCTCAG 1314
822 LeuGluGluIleGlnLeuValSerCysLeuSerAlaAsnAlaVally 838
1315 CTAGAACTCCTGATTTGGCATTTTACCGCTTACACATTAAATGCTCCACA 1364
838 sIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeuSerIle..... 852

1365 AAGTCCCTTCCAAACCTTCATTTCCCTTCAGGTTCTGAATCTCATCTACT 1414
853LeuAspLeuSerGlu..... 857
1415 GCTTCCTTGATACAGCAATTCAGACATCTTCTAGCAGGCTTACCAGTTCTC 1464
858 AsnTyrLeuGluLysAspLysnGluAlaLeuHisGluLeuIleAspAr 874
1465 CGGCATCTCACTTAAAGGAATCACTTCAAGATGGGACATATCAGAA 1514
874 gMetAsnValLeuGluGlnLeuThrAlaLeu.....MetLeuProTrpG 889
1515 GACCAACCTACTTCAAGACCGTGGAGCTTGGAGGTTCTGATTTGTCTCT 1564
889 LysAspValGlnGlySerLeuSerSerLeuLeuLysHisLeuGluGlu 905
1565 CTTGT.....GCTCTCTCTATATAGACCAGCAAGCATTCACACAGC 1605
906 ValProGlnLeuValLysLeuGlyLeuLysAsnTrpArgLeuThrAspTh 922
1606 TTGGGAAATATGAGCCATGTACACTTAAGCCACACAGCTGACATGCGA 1655
922 rGluIleArgIleLeuGlyAlaPhePheGlyLysAsnProLeuLysAsn 939
1656 CAGCATGTATCTCTTACGCCATCTTAAGGA..... 1686
939 heGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
1687 ..ATCTACCTCAATCTGGCTGCCAACAGCATTAAC 1719

THIS PAGE BLANK (USPTO)

CC Ischemia/reperfusion injury, excitotoxic brain damage, acute bacterial
XX meningitis and liver disease.

Sequence 3133 BP; 903 A; 691 C; 729 G; 810 T; 0 other;

alignment_scores:

Quality: 5281.00 Length: 1024
Ratio: 5.157 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AAS03945 ..

Sign seg 1/1 to: AAS03945 from: 1 to: 3133

```
1 MetAspPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyMe 17
|||||
36 ATGAATTTCAATAAGGACAAATAGCCGAGCCCTTATTCAAAGAAATGGGAA 85
|||||
17 TThrValIleLysGlnIleThrAspAspLeuPheValTrpAsnValLeuA 34
|||||
86 GACGTGTAAAGCAAAATCACAGATGACCTATTGTATGAAATGTTCTGA 135
|||||
34 snArgGluGluValAsnIleIleCysGluLysValGluGlnAspAla 50
|||||
136 ATCCGGAAGAAATCAATCATTTGCTCGAGAAAGTGAGACGATGCT 185
|||||
51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGluSerCysAs 67
|||||
186 GCTAGAGGATCATTCACATGATTTTGAAAAAGGTTTCAGACTCCTGTAA 235
|||||
67 nLeuPheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAsp 84
|||||
236 CCTCTTCTTAATCCCTTAAGAGATGGAACATATCCTATTTTCAGACT 285
|||||
84 euAsnGlyGlnSerLeuPheHisGlnTrpSerGluGlyAspLeuAspAsp 100
|||||
286 TGAATGGAACAAAGTCTTTTTCATCAGACATCAGAAGGAGACTTGGAGCAT 335
|||||
101 LeuAlaGlnAspLeuLysAspLeuTrpHisTrpProSerPheLeuAsnPh 117
|||||
336 TTGGCTCAGGATTTAAAGACTTGTACCATACCCATCTTTTCTGAACCT 385
|||||
117 eTrpProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerTrp 134
|||||
386 TTATCCCTTGGTGAAGATATGACATATTATTAACTTGAAGAACACACT 435
|||||
134 heThrGluProValLeuTrpArgLysAspGlnHisHisArgValGlu 150
|||||
436 TCACGAAACCTGTCTCTGGAGGAAGACCAACACCATCACCGCCTGGAG 485
|||||
151 GlnLeuTrpLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
|||||
486 CAGCTGACCTGAAATGGCCTCTCGAGGCTTTCAGAGCCCTCGATCAT 535
|||||
167 eGluGlyLysSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAla 184
|||||
536 TGAAGGGGAATCTGGCAAGGCAAGTCCACTCTGCTGACGCGCATTTGCCA 585
|||||
184 eLeuTrpGlySerGlyLysCysLysAlaLeuThrPheLysPheVal 200
|||||
586 TGCCTCTGGGGCTCGGAAAGTGAAGGCTTGACCAAGTTCAAATTTCTGC 635
|||||
201 PhePheLeuArgLeuSerArgAlaGlnGlyLeuPheGluTrpLeuCys 217
|||||
636 TTTCTTCTCCGCTTCAGCAGGCGCCAGGCTGACCTTTTGAACCCCTG 685
|||||
217 sAspGlnLeuLeuAspIleProGlyThrIleArgLysGlnTrpPheMet 234
|||||
686 TGATCAACTCTCGATATACCTGGCACAATCAGGAAGCAGACATTCATGG 735
|||||
234 IaMetLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTrp 250
```

```
|||||
736 CCATGCTGCTGAAGCTGGCGCAGAGGCTTCTTCTTCTGATGCGCTAC 785
|||||
251 AsnGluPheLysProGlnAsnCysProGluTrpIleGlnAlaLeuIleLysG 267
|||||
786 AATGAATTCAGAGCCCCAGAACTGCCAGAAATCGAAGCCCTGATTAAGGA 835
|||||
267 uAsnHisArgPheLysAsnMetValIleValIleThrThrThrGluCysL 284
|||||
836 AAACACCCGCTTCAAGAACATGATGTCATGTCACACATACCATGATGCC 885
|||||
284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
|||||
886 TGAGGACATACGGCAGAGTTGTGTCCTGACTGCTGAGGTGGGATATG 935
|||||
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLe 317
|||||
936 ACAGACAGACGCGCCAGGCTCTCATCCGAGAACTGCTGATCAAGAGACT 985
|||||
317 uAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
|||||
986 TGCTGAAGGCTTGTGCTCCAAATTCAGAAATCCAGAGGCTTGAAGATC 1035
|||||
334 euMetLysThrProLeuPheValIleThrCysAlaIleGlnMetGly 350
|||||
1036 TCATGAAGACCCCTCTCTTGTGTCATCACTGTGCATTCACATGAGTGG 1085
|||||
351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
|||||
1086 GAAAGTGAGTTCACCTCTCACACACAAACAGCTGTTCCTACCTTCTA 1135
|||||
367 rAspLeuLeuIleGlnLysAsnLysHisLysGlyValAlaIleAs 384
|||||
1136 TGATCTGTGTATACAGAAACAAACAAACATTAAGGTGGCTGCA 1185
|||||
384 eAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
|||||
1186 GTGACTTATTGGAGCCTGGACCACTGTGAGACCTGAGCTTGAGAGGT 1235
|||||
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
|||||
1236 GTGTTCTCCCAACAGTTGATTTCGAACCTGCAGATGTGTCACAGCTGAA 1285
|||||
417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysTrpThrAlaGln 434
|||||
1286 TGAGGATGTCTGCTGCAACACTGGGCTCTCTGTAATATACACTGAAA 1335
|||||
434 rGpHelysProLysTrpTrpPhePheHisLysSerPheGlnGluTrpThr 450
|||||
1336 GGTTCAGCCAAAGATTAATTTCTTTCACAAAGTCAATTCAGAGACTACACA 1385
|||||
451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
|||||
1386 GCAGGACAGACATCAGCAGATTATTTAGCTGTCAATGACCGCAGAGAGGT 1435
|||||
467 lThrLysGlyAsnGlyTrpLeuGlnLysMetValSerLysAspIleTr 484
|||||
1436 GACCAGGGAATGTGTTACTTGCAAGAAATGTGTTCCATTTCGAGACATTA 1485
|||||
484 hrSerThrTySerSerLeuLeuArgTrpTrpCysGlySerSerValGlu 500
|||||
1486 CATCACTTATAGCAGCTGCTCGGATACACTGTGGGTCTCTGTGTGGAA 1535
|||||
501 AlaThrArgAlaValMetLysHisLeuAlaIleValTrpGlnHisGlyCys 517
|||||
1536 GCCACAGAGGCTGTATTAAGACACTCCGACACATGATTAACACGGGTG 1585
|||||
517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnGluSerL 534
|||||
1586 CTTTCTCGAGCTTTCATCGCCAGAGAGCCCTCTGAGAGACGGAATCTT 1635
|||||
534 euGlnSerValLysAsnThrThrGluGlnGluIleLeuLysAlaIleAsn 550
|||||
```

1636 TGCAGAGTGTGAAAAACACCACTGAGCAGAGAAATTCGAAAAAGCCATAAC 1685
 551 ILeasnsrPheValGluCysGlyIleHisLeuTyrGlnGlnSerThrse 567
 1886 ATCAATTCCTTTGAGAGTGGCATCATTTATATCAAGAGATACATC 1735
 567 rlySerAlaLeuSerGlnGlnPheGlnAlaPhePheGlnGlyLysSerL 584
 1736 CAATATGAGCCCTGAGCAGATTTGAAGCTTCTTCAAGGTAAAGCT 1785
 584 eutyrlleasnsrGlyAsnIleProaspTyrLeupheaspPhePheGlu 600
 1786 TATATATCAATCAGGAGACATCCCGATTACTATTGACTTCTTGAA 1835
 601 HisLeuProaspCysAlaSerAlaLeuaspPheIleLysLeuaspPheTy 617
 1836 CATTTGCCCAATTTGTGCAGAGCTGTGACTGCTATTAACCTGGACTTTA 1885
 617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluaspThrGlyGly 634
 1886 TGGGGGAGCTATGGCTTCATGGGAAAGGCTGCAGAGACACAGGTGGA 1935
 634 leHisMetGlnGluAlaProGluThrTyrIleProSerArgAlaValSer 650
 1936 TCCACATGAGAAAGGCCCCAGAAACCTACATTCACAGAGGCTGTATCT 1985
 651 LeupheaspnsrTrpLysGlnIleupheArgThrLeuGlnValThrLeuAr 667
 1986 TTGTTCTTCACTGGAAGAGAGATTCAGAGCTGTGAGGTGCACACGCCG 2035
 667 gaSPheSerLysLeuasnLysGlnAspIleThrTyrLeuGlyLysIleP 684
 2036 GGAATTCAGACAGTTGAATAGCAGATATCACATATCTGGGAAAAATAT 2085
 684 heserSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
 2086 TCAGCTCTGCACAAAGCCCTGAGCTGCAAAATTAAGATGTGCTGTGTG 2135
 701 AlAGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
 2136 GCTGGAAGCCTCAGTTTGGTCTCAGACCTGTAAAGACATTTATTTCTCT 2185
 717 uMetValGluAlaSerProLeuThrIleGluaspGluArgHisIleThrS 734
 2186 CANGGTGGAAGCCAGTCCCTCACATAGAAATGAGAGCACATCATCAT 2235
 734 eTyValThrAsnLeuLysThrLeuSerIleHisaspLeuGlnAsnGlnArg 750
 2236 CTGTAAACAAACCTGAAACCTTGATGTATCATGACCTACAGAAATCAACG 2285
 751 LeuProGlyLysLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
 2286 CTGCCCGGTGCTGCTGACTGACAGCTTGGGTACTTGAGAAACCTTACAA 2335
 767 sLeuIleMetaspAsnIleLysMetAsnGlnGluaspAlaIleLysLeuA 784
 2336 GGCATATATGATTAACATTAAGATGAAGAAGATGCATATAAAGCTAG 2385
 784 laGluGlyLeuLysAsnLeuLysLysMetCysLeupheHisLeuThrHis 800
 2386 CTGAAGGCCCTGAAACCTGAAAGAGATGTGTATTATTCTTTGACCACAC 2435
 801 LeuSeraspIleGlyGlnGlyMetaspTyrIleValLysSerLeuSerse 817
 2436 TTGTTCTGACTTTGAGAGGGAATGATTAATAGTCAAGTCTCTGTCAAG 2485
 817 rGluProCysaspLeuGlnGluIleGlnLeuValSerCysCysLeuSerA 834
 2486 TGAACCTGTGACCTTGAAGAATTAATTAATGCTCCTGCTGTGCTG 2535
 834 laAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
 2536 CAAATGAGTGAATAATCTTAGCTCAGAAATCTTCACAATTTTGTCAAACTG 2585

851 SerIleLeuaspLeuSerGlnAsnTyrLeuGlnLysaspGlyAsnGlnAla 867
 2586 AGCATTCCTTGATTTATCAGAAAAATTAACCTGGAAGATGAAGATGAAGC 2635
 867 aleuHisGluLeuIleaspArgMetAsnValLeuGlnGlnLeuThrAla 884
 2636 TCTTCAATGAACATGATGACAGAGATGAACGCTGTAAGAACGCTACCGCAC 2685
 884 euMetLeuProTrpGlyCysaspValGlnGlySerLeuSerSerLeuLeu 900
 2686 TGATGCTGCCCTGGGCTGTGAGCTGCAAGGACGCTGAGACACGCTGTG 2735
 901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLysLeuasnTr 917
 2736 AAACATTTGGAGAGGATCCCAACATGCTCAAGCTTGAGGTGGAACATG 2785
 917 ParGLeuThrAspThrGluIleArgIleLeuGlnAlaPhePheGlyLysA 934
 2786 GAGACTCAGAGATACAGAGATTAAGATTTAGTGCATTTTGGAAAGC 2835
 934 snProLeuLysasnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
 2836 ACCCTCTGAAAAAATTCACAGACTTGAATTTGGCGGAAATCGTGTGAGC 2885
 951 SeraspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
 2886 AGTGATGATGATGCTTGCTTCATGGGTGTATTTGAAATCTTAAAGCAAT 2935
 967 uValPheaspPheSerThrLysGlnPheLeuProaspProAlaLeuV 984
 2936 AGTGTTTGTGACTTGTAGTACTTAAGAATTTCTACTGATCCAGCATTTG 2985
 984 alArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
 2986 TCAGAAACCTTACCAAGGTATTCAGAGTTAACTTTCTGCAAGAGCT 3035
 1001 ArgLeuValGlyTrpGlnPheaspAspAspAspLeuSerValIleThrG 1017
 3036 AGGCTGTGGGTGGCAATTTGATGATGATGATCTCAGTGTATTATACAG 3085
 1017 yAlaPheLysLeuValThrAla 1024
 3086 TGCTTTAAACTGTAATGCT 3107
 seq_name: /SIDS2/cgdata/geneseq/geneseqn/NA2001.DAT:AAH98254
 seq_documentation_block:
 ID AAH98254 standard: cDNA: 3545 BP.
 AC AAH98254;
 DT 12-OCT-2001 (first entry)
 DE Murine EST-derived coding sequence SEQ ID NO: 111.
 DE Human, sheep, pig, cow, fruit fly, yeast, hamster, macaque, horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 OS Mus musculus.
 XX WO200154477-A2.
 PN 02-AUG-2001.
 PD 25-JAN-2001; 2001WO-US02687.
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI: 2001-476164/51.
DR P-SDB: AAM23595.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
PS Claim 1: Page 250-251, 1275pp; English.

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, CC forensics, gene mapping, identification of mutations, to assess CC biodiversity and for nutritional purposes. The present sequence is a CC cDNA of the invention.

XX
XX Sequence 3545 BP; 1038 A; 755 C; 816 G; 936 T; 0 other:

alignment_scores:
Quality: 5259.00 Length: 1024
Ratio: 5.151 Gaps: 0
Percent Similarity: 99.707 Percent Identity: 99.707

alignment_block:
US-09-697-089-2 x AAM98254 ..
Align seg 1/1 to: AAM98254 from: 1 to: 3545

1 Metasnphellelysaspsnserrgalaaleuileglnatrgmetglyme 17
232 ATGAATTTTCATAAAGACAAATAGCCGACCTTATTCAAAGATGGGAAT 281
17 TThrValIlelysglnlethrAspAspleuPheValItrpAsnValIleua 34
282 GACTGTTTAAAGCAAATCACAGATGACCTATTGTATGAAATGTTCTGCA 331
34 snatrggluValasnlleleecyCysgluIysValgluAlaAspIa 50
332 ATCCGAGAGAGTAACATCATTTCTCGAGAAAGGTGAGACAGATGCT 381
51 AlaArgglyIlelleHisMetIleleuIysIysglYserglYserCysAs 67
382 GCTAGAGGGATCATTCACATGATTTTGAAAAGGGTTCAGAGTCTCTGTA 431
67 nleuPheleuIysSerleuIysgluItrpAsnItrpLeuPheGlnAspL 84
432 CCTCTTTTAAATCCCTTAAGAGAGTGAACACTATCCTATTTCAGGACT 481
84 euasnglyGlnSerleuPheHisGlnThrSerGluGlyAspleuAspAsp 100
482 TGAATGACAAAGTCTTTTCATCAGACATCAGAAAGAGACTTGAGCAT 531
101 LeuAlaGlnAspleuIysAspleuIyHisThrProSerPheleuAspPh 117
532 TTGGCTCGAGATTAAAGACTTGTACCATACCCCATCTTTCTGCAACT 581
117 eTyPProleuGluIysAspIleAspIleIlePheAsnleuIysSerThrP 134
582 TTATTCCTTGATGAGATGATGACATTTATTTACTTGAAGACACT 631
134 heThrgluProValleuItrpArgIysAspGlnHisHisArgValGlu 150
632 TCACAGAAACCTGTCTGTGAGGAAGACCAACACCATCACCGCTGGAG 681
151 GlnleuThrleuAsnglyleuIleuGlnAlaIleuGlnSerProCysIleI 167

682 CAGCTGACCCCTGAATGGCTCTCGAGGCTCTCAGAGCCCTGCATCAT 731
167 eGlnGlyIuSerGlyIysGlyIysSerThrleuGlnItrpGlnIleam 184
732 TGAAGGGGAATCTGGCAGAGCAAGTCACATCTGCTGACGCGCATTTGCCA 781
184 etleuItrpGlySerGlyIysCysIysAlaIleuThrIysPheIysPheVal 200
782 TGCTCTGGGGCTCCGGAAGATGCAAGGCTGTGACCAAGTCAATTTGCTC 831
201 PhePheleuArgleuSerArgAlaGlnGlyIyleuPheGluItrpIleu 217
832 TTTCTCTCCCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 881
217 SASpGlnleuIysAspIleProGlyThrIleArgIysGlnThrPheMet 234
882 TGATCAATCTCTGATATACCTGGCAGACATAGAGAGACAGACATTCATGG 931
234 IametleuIysleuArgGlnArgValIleuPheleuIysAspGlyTyr 250
932 CCATGCTGCTGAGAGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 981
251 AsnGluPheIysProGlnAsnGlyProGluIleGlnAlaIleuIysGln 267
982 AATGATTCAGAGCCCGAGAGCTGCCAGAAATCGAAGCCCTGATPAAAGGA 1031
267 uAsnHisArgPheIysAsnMetValIleValItrThrItrpGluCysL 284
1032 AAACCACCGCTTCAAGAAACATGTCATGTCACACATACCATCATGATGCC 1081
284 euArgHisIleArgGlnPheGlyAlaIleuThrArgIleValGlyAspMet 300
1082 TGAGGACATATGAGCAGATTTGTGCTGCTGCTGCTGAGTGGGAGATATG 1131
301 ThrGluAspSerAlaGlnAlaIleuIleArgGluValleuIleIysGluLe 317
1132 ACAGAGACAGCGCCAGCGCTCTCATCGAAGATGCTGATCAAGAGAGCT 1181
317 uAlaGlnGlyleuIleuGlnIleGlnIysSerArgCysleuArgAsnL 334
1182 TGCtGAAGGCTTTGTCCTCAAAATTCAGAAATCCAGGCTTGAGAGATC 1231
334 euMetIysThrProleuPheValIleThrCysAlaIleGlnMetCly 350
1232 TCATGAGAGACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1281
351 GluSerGluPheHisSerHisThrGlnThrleuPheHisItrpPheTy 367
1282 GAAAGTGAAGTTCACCTCTCACACACAAACAGCTGTTCCTACCTCTTA 1331
367 rAspleuIleuIleGlnIysAsnIysHisIysGlyValAlaIleAs 384
1332 TGATCTGTTGATACAGAAACAAACAAACAAACAAATAAAGGTGGCTGCA 1381
384 eRAspPheIleArgSerleuAspHisCysGlyAspleuAlaIleuGluGly 400
1382 GTGACTTCATTCGGAGGCTGGAGCAGCTGTGATACCTGCTGAGAGGCT 1431
401 ValPheSerHisIysPheAspPheGluIleuGlnAspValSerSerValAs 417
1432 GTGTTCTCCCAAGATTGATTTGAACTGCAAGATGTGTCACAGCGTAA 1481
417 nGluAspValleuIleuThrThrGlyleuIysCysIysTyrThrAlaGln 434
1482 TGAGAGATGCTCTGCTGACAACTGGGCTCTCTGTAATATACACTGAAA 1531
434 rGpHeIysProIysTyrIysPhePheHisIysSerPheGlnGluTyrThr 450
1532 GCTTCAAGCCAAAGATATATTTTTCAGAGTCATTCAGAGACTACACA 1581
451 AlaGlyArgIleuSerSerleuIleuThrSerHisGluProGluGluVal 467

1582 GCAGGACGACAGCTCAGAGTTTATTGACGTCTCATGACGACGACGAGCT 1631
 467 LThrlySGlyasnGlyIleuGlnIysMetValSerIleSeraspIleT 484
 1632 GACCAAGGGAGATGTTACTTGCAAGAAAATGTTCCATTTCGACATTA 1661
 484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
 1682 CATCCACTTATAGCAGCCTGCTCGGTACACCTCGGTGCTCTGTTGGA 1731
 501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyC 517
 1732 GCCACGAGGGCTGTTATGAAGCACCTCGCAGCAGTGTATCAACGCGCTG 1781
 517 sleuLeuGlyLeuSerIleAlaIysArgProLeuTyrArgGlnIleuSerL 534
 1782 CTTCTCGAGACTTTCATCGCCAAAGAGCCCTCTCGAGACAGAACTTT 1831
 534 euGlnSerValIysAsnThrThrGlnGlnIleuLeuLysAlaIleAsn 550
 1832 TGCAAAGTGTGAAAACACCACTGACGACAGAAATTCGAAAAGCATTAAC 1881
 551 IleasnSerPheValGluCysGlyIleHisLeuTyrGlnIleuSerThrSe 567
 1882 ATCAATTCCTTTGTAGAGTGTGCATCATTATATCAAGAGATACATC 1931
 567 rlySSerAlaIleuSerGlnIleuPheGluAlaPhePheGlnGlyIysSerL 584
 1932 CAAATTCAGCCCTGAGCCAGAAATTTGAAGCTTTCTTCAAGGTAAAGCT 1981
 584 euTyrIleasnSerGlyAsnIleProaspTyrLeuPheaspPheGlu 600
 1982 TATATATCAACTCGAGGACATCCCGATTACTTATTTGACTTCTTTGAA 2031
 601 HisLeuProasnCysAlaSerAlaLeuaspPheIleLysLeuaspPheTyr 617
 2032 CATTTGCCAATTGTGCAAGTGTCTGACTTCATTAACCTGGGCTTTTGA 2081
 617 rGlyAlaIleMetAlaSerTyrPglIuLysAlaIleGlnIleuSerThrGly 634
 2082 TGGGGAGCTATGGCTTCATGGAAGGCGCGAGAGACACAGGTGAA 2131
 634 leHisMetGlnGluAlaProGluThrTyrIleProSerArgAlaValSer 650
 2132 TCCACATGGAAGAGGCCCGCAAACTACATTCACAGAGGCGCTATCT 2181
 651 LeuPhePheasnTyrPlySGlnIleuPheArgThrLeuGluValThrLeuAr 667
 2182 TTGTTCTTCAACTGGAAGCAATTCAGACTCTGAGGCTCACACTCCG 2231
 667 gAsPheSerLysLeuasnIysGlnaspIleThrTyrLeuGlyLysIleP 684
 2232 GGATTTTCAGCAAGTTGAATAGCAAGATATCAGATATCTGGGGAAATAT 2281
 684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
 2282 TCAGCTCTGCCACAAAGCTCAGGCTGCAAAATAAAGATATGCTGCTGCTG 2331
 701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
 2332 GCTGGAAGCCCACTTGTGCTCCAGACCTGTAAAGAACTTTTCTCT 2381
 717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
 2382 CATGTGTGAAGCCAGTCCCTCACCATTAGAAAGAGAGGACACATCACAT 2431
 734 erValThrAsnLeuLysThrLeuSerIleHisAspIleGlnIleAsnGlnArg 750
 2432 CTGTAAACAAACCTGAAACCTGTAGTATTCTATGCTACAGAAATCAACGG 2481
 751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysasnLeuThrL 767
 2482 CTGCGGGGTGTGTGACTGACAGCTTGGTAACTTGAAGAACTTACAAA 2531

767 sleuIleMetAspAsnIleLysMetasnGlnIleuAspAlaIleLysLeuA 784
 2532 GCTCATATATGATTAACATAAAGATGAATGAAGAGATGCTATATAACTAG 2581
 784 laGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisIleuThrHis 800
 2582 CTGAAGGCCCTGAAAACCTGACAGAAGATGCTTATTTCATTGACCCAC 2631
 801 LeuSeraspIleGlyGlyMetaspTyrIleValIysSerLeuSerSe 817
 2632 TTGCTGTACATTGTGAGAGGGGAATGATTAATATGATCAAGTCTGTCAAG 2681
 817 rGluProCysAspLeuGlnIleuGlnLeuValSerCysCysLeuSerA 834
 2682 TGAACCTGTGACCTTGAAGAAATTCATTAATAGTCTCTGCTGTCTCT 2731
 834 laAsnAlaValIleLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
 2732 CAATGACAGTGAANAATCCTAGCTCAGATCTTCACATTTGGTCAAACTG 2781
 851 SerIleLeuasnLeuSerGluasnTyrLeuGlnLysaspGlyAsnGlnAl 867
 2782 AGCATCTTGATTATTCAGAAATTAACCTGCAAAAAGATGAATGAATGACG 2831
 867 aleuHisGluLeuIleaspArgMetasnValLeuGluGlnLeuThrAla 884
 2832 TCTTCATGAACCTGATGACAGATGAACGTGTGAACACCTCACCGCAC 2881
 884 euMetLeuProTyrPglIuLysAspValGlnGlySerLeuSerSerLeuLeu 900
 2882 TGATGCTGCCCTGGGCTGTGAGCTGCAAGGACCTGACACACTGCTG 2931
 901 LysHisLeuGlnIleuValProGlnLeuValLysLeuGlyLeuLysasnT 917
 2932 AAACATTTGGAGAGGCTCCCAACACTGTCMAAGCTGGCTTGA AAAAATG 2981
 917 parGluLeuThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLysA 934
 2982 GAGACTCACGATACAGATGAAATTTAGGTGCAATTTTGGAAAGA 3031
 934 snProLeuLysasnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
 3032 ACCCTCGAAAACCTCCAGCACTTGATTTGGCGGGAATCTGTGAGC 3081
 951 SerAspGlyTyrPleuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
 3082 AGTATGATGATGCGCTGCTTCATGAGGTGTATTTGAAGAACTTTAAGCAATT 3131
 967 uValPhePheAspPheSerThrLysGlnPheLeuProaspProAlaLeuV 984
 3132 AGTGTTTTGGACTTTAGTACTAAGAAATTTTACCTGATCCGCAATTTAG 3181
 984 alArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
 3182 TCAGAAAACCTTACCAAGTGTATCCAAGTTACTTTTCGCAAGAGCT 3231
 1001 ArgLeuValGlyTyrPglIleuPheaspAspAspLeuSerValIleThrGln 1017
 3232 AGGCTTGTGCTGGCAATTTGATGATGATGATGATGATGATGATGATGATG 3281
 1017 yAlaPheLysLeuValThrAla 1024
 3282 TGCTTTTAAACTAGTAACCTGCT 3303

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT: AAS03946
 seq_documentation_block:
 ID AAS03946 standard: DNA: 3615 BP.
 XX AAS03946;
 AC
 XX
 DT 12-SEP-2001 (first entry)

XX Human caspase recruitment domain 12 (CARD-12) genomic DNA.
DE
XX
KW Caspase recruitment domain: CARD-12; apoptosis; stress-related pathway;
KW cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KW Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;
KW aplastic anaemia; myocardial infarction; inflammatory disorder;
KW Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KW psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW tuberculosis; ischemic brain injury; hypoxic brain injury; des;
KW kidney ischemia; reperfusion injury; acute bacterial meningitis;
KW excitotoxic brain damage; liver disease.

Homo sapiens.

FH Key Location/Qualifiers
FT CDS 1..3615
FT /*tag= a
FT /product= "Human CARD-12"
PN MO200130971-A2.
XX
XX PD 03-MAY-2001.
XX
XX PE 26-OCT-2000; 2000WO-US29643.
XX
XX PR 27-OCT-1999; 99US-0161822.
XX
XX PA (MILL-) MILLENNIUM PHARM INC.
XX
XX PI Bertin J, Robison KE;
XX
XX DR WPI: 2001-308628/32.
XX P-PSDB: AAU02881.

XX Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease -
XX
XX

Disclosure; Fig 2; 93pp; English.

XX The sequence represents a genomic DNA which encodes the human caspase
XX recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
XX number of proteins that transmit signals that activate apoptosis and
XX inflammatory pathways in response to stress and other stimuli. Therefore,
XX CARD-12 and its corresponding nucleic acid may be used in treatment and
XX diagnosis of patients suffering from disorders associated with an
XX abnormal level (an increase or a decrease) of apoptotic cell death or
XX abnormal activity of stress-related pathways. The disorders include
XX cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
XX autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
XX neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
XX sclerosis), hematologic diseases (e.g. aplastic anemia, myocardial
XX infarction, stroke), inflammatory and immune system disorders (e.g.
XX Crohn's disease, insulin-dependent diabetes, contact dermatitis,
XX psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
XX lepromatous leprosy), ischemic and hypoxic brain injury, kidney
XX ischemia/reperfusion injury, excitotoxic brain damage, acute bacterial
XX meningitis and liver disease.

XX
SQ Sequence 3615 BP; 1041 A; 811 C; 845 G; 918 T; 0 other;

Alignment_scores:

Quality: 5214.00 Length: 1034
Ratio: 5.137 Gaps: 1
Percent Similarity: 98.162 Percent Identity: 98.066

alignment_block:
US-09-697-089-2 x AAS03946

Align seg 1/1 to: AAS03946 from: 1 to: 3615

1 MetAspPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlye 17
418 GTGAATTTCATATAAGGACATATAGCCGCTTATTTCATCAAGATGGAGAT 467
17 tThrValIleLysGlnIleThrAspAspLeuPheValTrpAsnValLeu 34
468 GACGTGTATAAAGCAATACAGATGACCTATTGTATGAAGATGTTCGA 517
34 snArgGluGluValAsnIleIleCysGluValGluGlnAspAla 50
518 ATCGCGAAGAAAGTAAACATCATTTCTGCGAGAGGTGACGAGATGCT 567
51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGlySerCysAs 67
568 GCTAAGAGGATCATTCACATGATTTTGAAGAAAGGTTCCAGACTCTGTAA 617
67 nLeuPheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAspL 84
618 CCTCTTCTTAAATCCCTTAAGAGAGTGAACATCTCTATTTCAGAGACT 667
84 euAsnGlyGlnSer..... 88
668 TGAATGACAAAGTTTGGAGACACAGAAATGGGCTCTTTAAACATC 717
89LeuPheHisGlnThrSerGluGlyAspLeuAs 99
718 ACCCTCTCTCTAATAGGCTCTTTTCATCAGACATAGAAAGGACTTGA 767
99 pAspLeuAlaGlnAspLeuLysAspLeuTrpHisThrProSerPheLeu 116
768 CGATTGGCTCAGATTTAAAGACTGTACCATACCCCATCTTTTCGA 817
116 snPheTrpProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSer 132
818 ACTTTTATCCCTTGGTGAGATATTGACATTATTTTAACTGTAAGC 867
133 ThrPheThrGluProValLeuTrpArgLysAspGlnHisHisArgVa 149
868 ACCCTCAGACGACGCTCCTCGAGAGAACCAACCATCACCGGCT 917
149 lGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysI 166
918 GGAGAGCGTACCCGTAATGGCTCCTCGAGGCTTTGAGAGCCCTGCA 967
166 lLeIleGluGlyLysSerGlyLysSerThrLeuLeuGlnArgIle 182
968 TCATTGAAGGGAATCTGCAAAAGGCAGAGTCCACTGCTGACGAGGAT 1017
183 AlaMetLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPh 199
1018 GCATAGCTCTGTGGGCTCGGAAAGTGCAAGGCTCGACCAAGTCAAT 1067
199 eValPhePheLeuArgLeuSerArgAlaGlnGlyIleLeuPheGluTrpL 216
1068 GCTCTTCTTCCTCCGCTCTCAGAGGAGGCCAGGGTGACTTTTGAACCC 1117
216 euCysAspGlnLeuLeuAspIleProGlyThrIleArgLysGlnThrPhe 232
1118 TCTGTGATCAACTCTCGATATACCTGGCACATATAGAGACAGACATTC 1167
233 MetAlaMetLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspG 249
1168 ATGGCCATGCTCTGAGCTGAGCGGAGAGGGTTTCTTCCTCTGATGG 1217
249 YTrpAsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIle 266
1218 CTACATATGAATTCAGCCCAAGACTGCCAGAAATTCAGAGCCCTGATAA 1267
266 YSGLuAsnHisArgPheLysAsnMetValIleValThrThrThrGlu 282
1268 AGGAAACACCGCTTCAAGAAATGTCATGCTGACCACTACCACTGAG 1317

283 CysLeuArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAs 299
|||||
1318 TGCCTGAGGACATACGCGAGTTGGGCTTGCCTGACTGCTGAGGTGGGGA 1367
299 PheThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLys 316
|||||
1368 TATGACAGAGACAGCGCCAGGCTCTCATCCGAGAAAGTCGTATCAAG 1417
316 IuLeuAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArg 332
|||||
1418 AGCTTGCTGAAGCTTGTGCTCCAAATTCAGAAATCCAGGTCTTGAGG 1467
333 AsnLeuMetLysThrProLeuPheValIleThrCysAlaIleGlnMe 349
|||||
1468 AATCTCATGAGACCCCTCTCTTGTGTCATCATCTTGTCATTCACAT 1517
349 tGlyGluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThr 366
|||||
1518 GGGTGAAGTGAAGTCCACTCCACACAAACACGCTTCCATACCT 1567
366 heTyrrAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAla 382
|||||
1568 TCTATGATCTGTGATACAGAAAAACAAACAAACATAAAGCTGTGGCT 1617
383 AlaSerAspPheIleArgSerIleuAspHisCysGlyAspLeuAlaLeu 399
|||||
1618 GCATGTCATCTTCATTCGAGCTGAGACCTGAGACCTGAGACCTGAG 1667
399 uGlyValPheSerHisLysPheAspPheGluLeuGlnAspValSerSer 416
|||||
1668 GGGTGTGCTTCCACAAAGTTGATTCGAACTGACGAGATGTGCCAGC 1717
416 alaGlnGluAspValLeuLeuThrThrGlyLeuLeuLysLysThrAla 432
|||||
1718 TGAATGAGATGTCCTGCTGACACAGCGCTCTGTAAATATACAGCT 1767
433 GlnArgPheLysProLysTyrLysPhePheHisLysSerPheGlnLys 449
|||||
1768 CAAGGTTCAGCCCAAGTAAATCTTTACACAGTCATTCACAGAGAT 1817
449 rThrAlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGlu 466
|||||
1818 CACAGCAGCAGCAAGACTCAGAGTTTATGAGCTCATGAGCCAGAG 1867
466 IuValThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAsp 482
|||||
1868 AGGTGACCAAGGGGAATGTTACTTCAGAAATGGTTCCATTTGGGAC 1917
483 IleThrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSer 499
|||||
1918 ATTACATCCCTTATAGCAGCTGCTCGGTACACCTGCGGTGATCTGT 1967
499 IguAlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHis 516
|||||
1968 GGAAGCCACAGCGGCTGTTATGAAGCAGCTCGACAGTGTATCAAC 2017
516 LysCysLeuLeuGlyLeuSerIleAlaLysArgProLeuThrPargGln 532
|||||
2018 GGTGCTTCGAGCTTCCATGCGCAAGAGGCTCTCGAGACAGGA 2067
533 SerLeuGlnSerValLysAsnThrThrGlnGlnGlnIleLeuLysAla 549
|||||
2068 TCTTTCGCAAGTGTGA AAAACACACTGAGCAAGAAATTCGAAACCAT 2117
549 eAsnIleAsnSerPheValGluCysGlyIleHisLeuTyrGlnLysSer 566
|||||
2118 AAACATCAATTCCTTGTAGAGTGTGCAATTCATTTATCAAGAGAGTA 2167
566 hSerLysSerAlaLeuSerGlnGluPheGlnAlaPhePheGlnGlyLys 582
|||||
2168 CATCAAAATCAGCCGTGAGCCAGAAATTTGAAGCTTTCTTCAAGGTAA 2217

583 SerLeuTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPhe 599
|||||
2218 AGCTTATATATCACTACAGGAACATCCCGATTTACTTATTTGACTTCT 2267
599 eGluHisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAsp 616
|||||
2268 TGAACATTTGGCCAAATTTGTCAAGTCCCTGAGCTCATTTAAACGTGACT 2317
616 heTyrrGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGly 632
|||||
2318 TTTATGGGGAGCTATGCTTCATGGGAAAGCTGACAAACACAGCT 2367
633 GlyIleHisMetGlnGluAlaProGluThrTyrIleProSerArgAla 649
|||||
2368 GGAATCCACATGAGAGAGGCCCCAGAAACCTTACATTCACAGAGGCTGT 2417
649 IserLeuPhePheAsnTrpLysGlnIuPheArgThrLeuGluValThr 666
|||||
2418 ATCTTGTTCTTCAACTGAGAGAGGAATTCAGAGCTGAGAGTCCACAC 2467
666 euArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLys 682
|||||
2468 TCCGGGATTTCAAGCAAGTTGAATGAACAGATATCAGATATCTGGGAAA 2517
683 IlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAla 699
|||||
2518 AATTTACGCTTGCCACAGACCTCAGCTGCAGAAATGAAGATGTGCTGG 2567
699 yValAlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleLys 716
|||||
2568 TGTGGCTGAGAGCTCAGTTTGTCTCAGCACCTGATGAAGACATTTAT 2617
716 eLeuMetValGluAlaSerProLeuThrIleGluAspGluArgHisIle 732
|||||
2618 CTCTCATGTGTGAAGCCAGTCCCTCACCATGAAGATGAGAGGACAC 2667
733 ThrSerValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsn 749
|||||
2668 ACATCTGTAAACAACCTGAAAACCTTGAGTATTCATGACCTACAGATCA 2717
749 nArgLeuProGlyGlyLeuThrAspSerLeuGlnLysLeuLysAsnLeu 766
|||||
2718 ACGGTCGCGGGTGTGCTGACTGACAGCTTGGTGAACCTTGAAGAACCT 2767
766 hTrpLysLeuIleMetAspAsnIleLysMetAsnGluGlnAspAlaIleLys 782
|||||
2768 CAAAGCTCATTAATGATGAATGAATGAATGAAGAAATGCTATAAAA 2817
783 LeuAlaGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeu 799
|||||
2818 CTAGCTGAAGGCTGAAAAAAGCTGAAAGAAAGATGTGTTTATTTCAATTTGAC 2867
799 rHisLeuSerAspIleGlyGluGlyMetAspTyrIleValLysSerLeu 816
|||||
2868 CCACCTGTCTGACATTTGAGAGGAATGATTAACATAGCAAGTCTCTGT 2917
816 eSerGluProCysAspLeuGluGlnIleGlnLeuValSerCysLeu 832
|||||
2918 CAAGTGAACCTGTGACCTTGAAGAAATTCATTAAGTCCGCTGCTTG 2967
833 SerAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuVal 849
|||||
2968 TCTGCAAAATGCAAGTAATCTTACTGTCGATCTTCACAAATTTGTGCA 3017
849 sLeuSerIleLeuAspLeuSerGluAsnTyrLeuGlnLysAspGlyAsn 866
|||||
3018 ACTGAGCATTTCTGATTTATCAGAAAAATTAACCTGAAAAAATGAGAAATG 3067
866 IuAlaLeuHisGluLeuIleAspArgMetAsnValLeuGlnIleLeuThr 882
|||||
3068 AAGCTCTTATGAACTGATCGACAGAGATGAACGTCTGAAACAGCTCAC 3117
883 AlaLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerIle 899

```

3118 GCACGATGCTGCGCCCTGGGGCTGGAGCTGCAGGCAAGCGCTGAGCAGCGCT 3167
899 uleuLysHISleuGluGluValProGlnLeuValLysLeuGlyLeuLysA 916
1168 GTTGAACAATTGGAGAGAGGTCCACAACTGTCMAAGCTTGGGTGAAAA 3217
916 sntPrpArgLeuThrAspThrGluIleArgIleLeuGlyAlaIlePhePheGly 932
3218 ACTGGAGACTCACAGATACAGAAATTTAGATTTTAGGTGCATTTTTCGGA 3267
933 LysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgVal 949
3268 AAGAACCCCTCTGAAAAAATCTCCAGAGCTTGAATTTGGCGGGAATTCGNGT 3317
949 LserSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysG 966
3318 GAGCAGTGTATGATGGCTTCCTGCCTTCAAGGGTGATTTTGGAGATTTTAAAG 3367
966 ILeuValAlaPhePheAspPheSerThrLysGluPheLeuProAspProAla 982
3368 AATTAGTGTATTTTGGACTTTAGTACTTAAGAATTTCTACCTGCATCCAGCA 3417
983 ILeuValArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnG 999
3418 TTAGTCAGAAAACTTAGCCAAAGTTATATCCAGCTTAATCTTTCGCAGA 3467
999 uAlaArgLeuValGlyTrpGlnPheAspAspAspAspLeuSerValIleT 1016
3468 ACCTAGACCTTGTGGGTGGCCAAATTTGATGATGATGATCAGTGTATTA 3517
1016 hr 1016
11 3518 CA 3519
seq_name: /SIDS2/ycgdata/geneseq/geneseqn/NA2001.DAT.AAH9551

```

seq. documentation block:	seq. documentation block:
ID	ID
AAH99581 standard; cDNA; 2950 BP.	AAH99581 standard; cDNA; 2950 BP.
AAH99581;	AAH99581;
16-OCT-2001 (first entry)	16-OCT-2001 (first entry)
Human protein encoding cDNA sequence SEQ ID NO:416.	Human protein encoding cDNA sequence SEQ ID NO:416.
Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antihistaminic; antihistaminic; antihistaminic; immunosuppressive; antihistaminic; endocrine; cardiac; central nervous system; vitamin; anti-HIV; fungicide; antimutagen; cardiovascular; antineoplastic; anaemia; antileukemic; haemostatic; vulnery; antileukemic; osteopathic; eczema; dermatological; antileukemic; antileukemic; antidiabetic; cytostatic; neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antineoplastic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.	Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antihistaminic; antihistaminic; antihistaminic; immunosuppressive; antihistaminic; endocrine; cardiac; central nervous system; vitamin; anti-HIV; fungicide; antimutagen; cardiovascular; antineoplastic; anaemia; antileukemic; haemostatic; vulnery; antileukemic; osteopathic; eczema; dermatological; antileukemic; antileukemic; antidiabetic; cytostatic; neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antineoplastic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.
Homo sapiens.	Homo sapiens.
MO200153455-A2.	MO200153455-A2.
26-JUL-2001.	26-JUL-2001.
22-DEC-2000; 2000MO-US35017.	22-DEC-2000; 2000MO-US35017.
23-DEC-1999; 99US-0471275.	23-DEC-1999; 99US-0471275.
21-JAN-2000; 2000US-0488725.	21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0552317.	25-APR-2000; 2000US-0552317.

PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Dzmanac RT;
 XX
 DR WPI: 2001-457603/49.
 XX
 P-PSDB; AAM25640.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 XX
 PS Claim 1; Page 511-512; 1217pp; English.
 XX
 CC AAh99166 to AAh99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antitumagen;
 CC cardiovascular; antihaemic; antiagregant; haemostatic; vulnerary;
 CC antitumor; osteoplastic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anapylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX
 SQ Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other;

```

alignment_scores:      Quality: 3266.00      Length: 642
                       Ratio: 5.111      Gaps: 0
Percent Similarity: 99.533      Percent Identity: 99.377

alignment_block:
US-09-697-089-2 x AAH9581 ..

Align seg 1/1 to: AAH9581 from: 1 to: 2950

383 AAlaSerAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGI 3999
|||||
|||||
2 GCAAGTGACTTCATTCGGAGGCTGGACCACTGGTGAATCTCTCGGA 51

399 uGlyAlaPheSerHisLysPheAspPheGluLeuGlnAspValSerSerV 416
|||||
|||||
52 GGGTGTGTTCTCCCAAGTTTGATTCGAACTGCAGAGTGTGCCAGC 1010

416 aAlaGlnLysPValLeuLeuThrThrGlyLeuLeuCysLysTYrThrAla 432
|||||
|||||
102 TGAATGAGAGATGTCCTCGTGCACAACTGGCGCTCTCTGTAATATACGACT 1511

433 GlnATGPhelysProLysTYrLysPhePheHisLysSerPheGlnGluTY 449
|||||
|||||
152 CAAAGGTTCAAGCCAAAGTATTAATTTCTTTCACAGTCATTCACAGAGTA 2010

449 rThrAlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluG 466
|||||
|||||
202 CACACAGAGACCAAGACACACAGAGTTTATTCAGCTGCATGAGCCACAGC 2511

466 lValAlaThrLysLysGlnGlyTYrLeuGlnLysMetValSerIleSerAsp 482
|||||
|||||
252 AGGTGACCAAGGGGAAATGTTACTTGCAGAAATGGTTTCCATTTCCGAC 3010

483 lIleTrnSerThyTySerSerLeuArgTrnThyCysGlySerSerVal 499

```

|||||
302 ATTACATCCACTTATAGACAGCTGCTCGGTACACCTGTGGTCATCTGT 351
499 IGLUALATHATGAlaValMetLysHISLeuAlaAlaValTyrGlnHISG 516
352 GGAAGGCACAGGGCTGTTATGAAGACCTCCGACAGTGTATCAACACG 401
516 LysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrrArgGlnGlu 532
402 GCGGCTTCTCGGACTTTCATGCGCAAGAGGCTCTCTGGAGACAGAA 451
533 SerLeuGlnSerValLysAnthrTrhGlnGlnLileuLysAlaI 549
452 TCTTGCAAAAGTGTGAAAAACACCTGAGCAAGAAATTCGAAACCAT 501
549 eAnIleAnSerPheValGluCysGlyIleHISLeuTyrGlnGlnSer 566
502 AAACATCAATTCCTTGTAGAGTGTGCATCCATTTATATCAAGACAGTA 551
566 hrSerLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLys 582
552 CATCCAAATCAGCCCTGAGCCAAAGAAATTTGAAGCTTTCTTTCAAGTAA 601
583 SerLeuTyrIleAnSerGlyAsnIleProAspTyrLeuPheAspPheP 599
602 AGCTTATATATCAACTCAGGAGAACATCCCGATTACTTATTTGACTTCT 651
599 eGluHISLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAsp 616
652 TGAACATTTGGCCCAATTTGTCAAGTCTCTGGACTTCATTTAACTGGGCT 701
616 heTyrGlnGlyAlaMetAlaSerTrpGlnLysAlaIleGluAspTrhGly 632
702 TTTATGGGGAGCTATGGCTTCTATGGGAAAGGCTGCAGAGACAGAGGT 751
633 GlyIleHISMetGlnLualProGluTrhTyrIleProSerArgAlaVal 649
752 GGAATCCACATGAGAGAGGCCCAAGAACCTTACATTCGCCAGCAGGCTGT 801
649 LserLeuPhePheAsnTrpLysGlnLupheArgTrhLeuGluValTrhL 666
802 ATCTTTGTCTTCACTGGAAGCAGAAATTCAGGACTCTGGAGGTCAAC 851
666 eArgAspPheSerLysLeuAsnLysGlnAspIleTrhTyrLeuGlyLys 682
852 TCCGGGATTTCAAGCAAGTTGAATAGCAACATATTCAGATATCTGGGGAAA 901
683 IlePheSerSerAlaTrhSerLeuArgLeuGlnIleLysArgCysAlaI 699
902 ATATTCAGCTCTGCCACAAGCCTCAGGCTGCAGAAATTAAGAGATGCTGTG 951
699 yValAlaGlySerLeuSerLeuValLeuSerTrhCysLysAsnIleTyrS 716
952 TGTGGCTGGAGGCTCAGTTGTCTCTCAGCACCTGTAAAGAACATTTATT 1001
716 erLeuMetValGluAlaSerProLeuTrhIleGluAspGluArgHISile 732
1002 CTCTCATGTGGGAGAGCCAGTCCCTCACCATTAGAACATGAGAGGCCATTC 1051
733 TrhSerValTrhAsnLeuLysTrhLeuSerIleHISAspLeuGlnAsnG 749
1052 ACATCTGTAAACAACCTGAAACCTTGAGATATTCATGACCTACAGAAATCA 1101
749 narGluProGlyGlyLeuTrhAspSerLeuGlyAsnLeuLysAsnLeu 766
1102 ACGGCTGCCGGGTGTCTGCTGACGCTTGGGTAACTTGAAACACTTA 1151
766 hrLysLeuIleMetAspAsnIleLysMetAsnGlnGluAspAlaIleLys 782
1152 CAAAGCTCATATGATGAATACATTAAGATGATGAACAGACTGCTATTAATA 1201
783 LeuAlaGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHISLeuTh 799
|||||

1202 CTAGCTGAAGGCTGTGAAAAACCTGAAGAAGATGTGTTATTTTCATTTGAC 1251
799 rHISLeuSerAspIleGlyGluLysMetAspTyrIleValLysSerLeuS 816
1252 CCAGTGTCTGACATTGAGAGAGGAATGATTACAAATGCAACTGCTGTG 1301
816 erSerGluProCysAspLeuGluGluIleGlnLeuValSerCysCysLeu 832
1302 CAAGTGAACCCCTGTGACCTTTGAAGAAATTCATTTAGTCTCTCTGCTTG 1351
833 SerAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHISAsnLeuValLys 849
1352 TCTGCAAAATGTGAGTGAATAATCTAGCTCAGAAATCTTCACAAATTTGGTCA 1401
849 sLeuSerIleLeuAspLeuSerGluAsnTyrLeuGluLysAspGlyLysArg 866
1402 ACTGACATTCCTTGTGATTTATCAGAAAAATTAACCTGAAAAAGAGGAAATG 1451
866 lualLeuHISGluLeuIleAspArgMetAsnValLeuGlnGlnLeuTrh 882
1452 AAGCTTCTCATGACTGATGACAGAGATGACAGTGTCTAGAACAGCTCACC 1501
883 AlaLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLe 899
1502 GCACATGATGCTGCCCTGGGGCTGTGACGTGCAGAGGAGCAGCCTGAGACCT 1551
899 uLeuLysHISLeuGluGluValProGlnLeuValLysLeuGlyLeuLysA 916
1552 GTTGAACAATTTGGAGAGGTCCACACACTCGCAAGCTTGGGTTGAAAA 1601
916 sntPrpArgLeuTrhAspTrhGluIleArgIleLeuGlnAlaPhePheGly 932
1602 ACTGGAGACTCACAGATACAGAGATTAAGATTTTACGTGATTTTGGCA 1651
933 LysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgVa 949
1652 AAGAACCTCTGTGAAAAAATTCAGCAGTGAATTTGGCGGAAATGCTGTG 1701
949 LserSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysG 966
1702 GAGCAGATGATGAGTGGCTTCTCATGGGTGATTTGAGAAATCTTAAC 1751
966 lneValPhePheAspPheSerTrhLysGlnPheLeuProAspProAla 982
1752 AATTAGTGTTTTGTGACTTTAGACTTAAGAAATTTCTACTGTATCCAGCA 1801
983 LeuValArgLysLeuSerGlnValLeuSerLysLeuTrhPheLeuGlnG 999
1802 TTAGTCAGAAACTTAGCCAAAGTGTATCCAAAGTTAACTTTCTGCAAGA 1851
999 uAlaArgLeuValGlyTrpGlnPheAspAspAspAspLeuSerValIle 1016
1852 AGCTAGGCTTGTGGGTGGCAATTTGATGATGATGATGATGATGATGATTA 1901
1016 hrGluAlaPheLysLeuValTrhAla 1024
1902 CAGGTGCTTTTAACATAGTAACGTCT 1927
seq_name: /SID52/gcdata/geneseq/geneseq/NA2001.DAT:AAH34171
seq_documentation_block:
ID AAH34171 standard; cDNA: 2735 BP.
XX
AC AAH34171;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1253.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.

```

XX  WO200122920-A2.
XX
XX  05-APR-2001.
XX
XX  28-SEP-2000: 2000MO-US26524.
XX
XX  29-SEP-1999: 99US-0157137.
XX  03-NOV-1999: 99US-0163280.
XX
XX  (HUMA-) HUMAN GENOME SCI INC.
XX
XX  Ruben SM, Barash SC, Birse CE, Rosen CA:
XX
XX  WPI: 2001-235357/24.
XX  P-PSDB: AAG74766.
XX
XX  Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX  useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX  Claim 1: Page 3017: 9803pp: English.
XX
XX  AAH37943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX  cancer-associated nucleic acid molecules (N) and proteins (P), where
XX  the proteins are collectively known as colon cancer antigens. The colon
XX  cancer antigens have cytostatic activity and can be used in gene
XX  therapy and vaccine production. N and P may be used in the prevention,
XX  diagnosis and treatment of diseases associated with inappropriate P
XX  expression. For example, N and P may be used to treat disorders
XX  associated with decreased expression by rectifying mutations or deletions
XX  in a patient's genome that affect the activity of P by expressing
XX  inactive proteins or to supplement the patient's own production of P.
XX  Additionally, N may be used to produce the colon cancer-associated Ps,
XX  by inserting the nucleic acids into a host cell and culturing the cell
XX  to express the proteins. N and P can be used in the prevention, diagnosis
XX  and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX  and AAB77788 represent sequences used in the exemplification of the
XX  present invention.
XX  N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX  missing at time of publication, meaning no sequences are present for
XX  SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX  Sequence 2735 BP; 791 A; 555 C; 512 G; 876 T; 1 other:

alignment_scores:
    Quality: 921.00      Length: 180
    Ratio: 5.117        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AAH34171/rev ..

Align seg 1/1 to reverse of: AAH34171 from: 1 to: 2735

845 HisaAnleuValLysLeuSerIleLeuAspLeuSerGluAsnTyrLeuGI 861
|||||
2727 CACAATTGGTGCACAACTGACATCTTGATTTATCAGAAATTAACCTGGA 2678
|||||
861 uLyAspGlyAsnGluAlaLeuHisGluLeuIleAspArgMetAsnValL 878
|||||
2677 AAAAGATGGAAATGAAAGCTCTTCATGAACTGATCGACAGATGAACGTGC 2628
|||||
878 euGluGlnLeuThrAlaLeuMetLeuProTfPGlyCyAspValGlnGly 894
|||||
2627 TAGAACAGCTACCGCACTGATGCTGCGCTGGGGCTGTGACGTGCAAGCC 2578
|||||
895 SerLeuSerSerLeuLeuLysHisLeuGlnLysValProGlnLeuVally 911
|||||
2577 AGCTGAGACACCTGTGTGAACATTTGGAGAGGTCCACAACTGCTCAA 2528
|||||
911 sLeuGlyLeuLysAsnTfPAArgLeuThrAspThrGluLeuArgIleLeuG 928
|||||

```

```

2527 GCTTGGGTTGAAAAAAGCTGAGACTCAGATACAGATTAGAAATTTAG 2478
928 lYAlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeu 944
|||||
2477 GTGCATTTTGTGGAGAGAACCTCTGAAAAACCTCCACAGATTGAAATTTG 2428
|||||
945 AlaGlyAsnArgValSerSerAspGlyTfRpleuAlaPheMetGlyValPh 961
|||||
2427 GCGGGAATTCGTGTGAGCAGATGATGATGCTTGCTTCATGGGGTGTATT 2378
|||||
961 eGluAsnLeuLysGlnLeuValPhePheAspPheSerThrLysGluPheL 978
|||||
2377 TGAGATCTTAAAGCAATATGATTTTGAAGTTAGTCTTAAAGCAATTTT 2328
|||||
978 euProAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLeu 994
|||||
2327 TACCTGATCCAGCATTTAGTCAGAAAACCTTACCCAGTGTATCCAGTTA 2278
|||||
995 ThrPheLeuGlnGluAlaArgLeuValGlyTfRpleuPheAspAspAsp 1011
|||||
2277 ACTTTCTGCAAGAGCTAGGCTTGTGTGGTGGCAATTTGATGATGATGA 2228
|||||
1011 PleuSerValIleThrGlyAlaPheLysLeuValThrAla 1024
|||||
2227 TCTCAGTGTATTACAGGTGCTTTTAAACTAGTAAGTGTCT 2188
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX58000
seq_documentation_block:
ID AAX58000 standard: cDNA to mRNA; 5984 BP.
XX
XX  AAX58000:
XX
XX  20-JUL-1999 (first entry)
XX
XX  Gonadotropic hormone coding sequence.
XX
XX  Gonadotropic hormone; excessive ovulation animal; transgenic animal;
XX  totipotent cell; somatic cell chromosome; ds.
XX
XX  Homo sapiens.
XX
XX  JP11113444-A.
XX
XX  27-APR-1999.
XX
XX  14-OCT-1997: 97JP-0280830.
XX
XX  14-OCT-1997: 97JP-0280830.
XX
XX  (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX  (SAKA/) SAKAI H.
XX
XX  WPI: 1999-320709/27.
XX  P-PSDB: AAY14079.
XX
XX  An excessive ovulation animal - useful for improving the
XX  productivity of animals
XX
XX  Claim 1: Page 7-9; 18pp; Japanese.
XX
XX  This sequence encodes a gonadotropic hormone.
XX  The invention relates to an excessive ovulation animal, which is a
XX  transgenic animal with a totipotent cell containing a DNA fragment
XX  containing a promoter sequence and a gonadotropic hormone coding
XX  sequence. The DNA fragment is in the somatic cell chromosome. The
XX  excessive ovulation animal is useful for improving the productivity of
XX  animals. The method can improve the productivity of a useful animal.
XX
XX  Sequence 5984 BP; 1777 A; 1267 C; 1267 G; 1673 T; 0 other:

alignment_scores:

```

Quality: 653.00 Length: 1108
Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736

Alignment_block:

US-09-697-089-2 x AAX58000 ..

Align seg 1/1 to: AAX58000 from: 1 to: 5984

69 PheLeuLysSerLeuLysGluTrpAsnTyrProLeuPheGlnAspLeu 85
|||||:.....:|||||:.....: |||||:
1321 TTTCTCCAAATATGAAAGTCC...TCGCGGAAGTGACTCCAGACCTTCA 1367
85 ngLgln.....SerLeuPheHisGlnThrSerGluLysAsp 98
:|||||:.....: |||||:.....: |||||:
1368 GACCGCGGGAAGTCTTGTGATTTACTGGAACACAGTGAAGCAATC 1417
98 euAspAsp.....:.....:.....: 100
|||||:
1418 TTGAGATTCATATAGCAGTTGTCTATAGTCCAGAAATGCACAGGGT 1467
101LeuAlaGlnAspLeuLysAs 107
1468 GAAGCCCAAGTGTTCAGAGGCAAGAAATGATGACAGCTGAGAGC 1517
107 PLeuTyrHisThrProSerPheLeuAsnPro..... 117
|||: |||||:.....: |||||:
1518 AGCTTATACAGAGCGGCAATTCGCCCAATGCTTTCCTGATATCTCT 1567
118TyrProLeuGluLysAspLeuAspLeuPhe 128
:|||||:.....: |||||:.....: |||||:
1568 CCGATCTGGCCAGCAGCACTGTGGGCTGTGATCTGTCTATT..... 1611
129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGln 145
:.....: |||||:
1612 ...GCTTCAAAACACATCAGCAAACTGTG..... 1638
145 SHISHisArgValGluGlnLeuThrLeuAsnGlyLeuGlnAlaLeu 162
||| |||||:.....: |||||:
1639CAGAACCTCTGCTGCTGCTGAGTCTTTGGCACTGTA 1678
162 InSerProCysIleIleGluGlyLysSerGlyLysGlyLysSerThrLeu 178
:|||||:.....: |||||:.....: |||||:
1679 ACCTCTGATGCTGTGAGGAGTGAAGCTGGAAGTGAAGACGAGCTTC 1728
179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeu 195
|||||:.....: |||||:.....: |||||:
1729 CTGAGAAATATACCTTTCTGTGGCACTGATGCTGCTCCCTGTTAA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
:|||||:.....: |||||:.....: |||||:
1779 CAGGTCACAGCTGTTTCTACTCTCCCTTACTTCCACCAACAGCAGAG 1828
210 LysIleLysPheGluThrLeuCysAspGlnLeuLysAspIleProGlyThr 226
|||||:.....: |||||:.....: |||||:
1829 AGGCGCTGGCCAGTATCTGTGACCAAGCTCCTAGAAAGAAAGATCT 1878
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArg 243
:|||||:.....: |||||:.....: |||||:
1879 GTTACTGAAATGTCATGAGCAATTTATCCAGACGTTAAAGATCAGGT 1928
243 IleuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258
|||||:.....: |||||:.....: |||||:
1929 CTTATTCCTTTTATGATGACTACAAAGAAATATGTCAATCCCTCAA... 1974
258 ysProGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
||| |||||:.....: |||||:
1975GTCAATAGGAAACCTGATTCAAAACCACTTATCCCGAGCTGC 2019
275 ValIleValThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
:|||||:.....: |||||:.....: |||||:
2020 CTTATGATTCCTGCTCCGTACAAACAGGCGCAGGACATCCGCGATACCT 2069

291 ValAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
|||||:.....: |||||:.....: |||||:
2070 AGAGACCATTCCTAGACATCAAAACATTTCCCTTTTAAATACGTCTGTA 2119
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
:|||||:.....: |||||:.....: |||||:
2120 TATTACGGAAGCTCTTTTACATATAATATGACTCGTCCGCAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProle 339
:|||||:.....: |||||:.....: |||||:
2170 GTTTACTTTGGAAAGAACCAAGTTTGCAAGATACAGAAACCTCCTCT 2219
339 uphValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
|||||:.....: |||||:.....: |||||:
2220 CTTTGGGCGCGCATCTGTCTCATCTGTTTGCAGTATCCTTTGACCCAT 2269
356 eHisThrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
||| |||||:.....: |||||:.....: |||||:
2270 CCTTGATGATGTGGCGTGTTCAGTCCATATGGAACGCTTTCCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheLeuArg 388
|||||:.....: |||||:.....: |||||:
2320 AGGAACAAA.....GCGACAGCTGAATTCCTCAAGC 2351
389 SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405
:|||||:.....: |||||:.....: |||||:
2352 AACTGTCTCCTCTGTGTAGCTGGCTGCTGGAAGGTTTTTTCATGTT 2401
405 ysPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
|||||:.....: |||||:.....: |||||:
2402 GCTTATGATTTATGATGATGATCTCGCAGAACAGGAGGTGATGAAGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
|||||:.....: |||||:.....: |||||:
2452 GAAGATCTAACATGTGCTGATGACCAAAATTTACGCCCGAGCTAAG 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGly 453
||| |||||:.....: |||||:.....: |||||:
2502 ACCATTCCTACCGGTTTTTAAGCTCTGCTCCCAAGAAATTTCTGGCGGA 2551
453 rArgLeuSerSerLeuLeuThrSerHisGluProGluValThrLys 469
|||||:.....: |||||:.....: |||||:
2552 TGAGCGCTGATTAACCTCGATTCAGATAGCGAGACATCAAGATTTG 2601
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSer 486
||| |||||:.....: |||||:.....: |||||:
2602 GAGCTGATATCTTGAACAAATCAACTCACCCATGATGACTGAAGCGC 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
:|||||:.....: |||||:.....: |||||:
2652 CTACAAACATTTTGTGACATATGTC.....TCCAGCTCCCTTCACAA 2695
503 rGluAla.....ValMetLysHisLeuAlaValIleGlnHisGly 516
:|||||:.....: |||||:.....: |||||:
2696 AACACAGGCGCCAAATTTGTCTCATTTGTCTCATTTAGTGGAATAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnLys 533
:|||||:.....: |||||:.....: |||||:
2743AAAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
|||||:.....: |||||:.....: |||||:
2751 ATTGGAATATATATGAAATGATGACTACTTAAGCAGCAGCCAGAA 2800
540 ..ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551
:|||||:.....: |||||:.....: |||||:
2801 TTTCACGTGAGATGACAGTACTTAGGGGATTTGGCAAAATTTGTCCAAA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
||| |||||:.....: |||||:.....: |||||:
2851 GCTTACTTTTCAATGCTTTCAGAACATTTTACTGCTTCCGCTGAAGC 2900
562 ...TyrGlnGlnSerThrSerLysSerAlaLeuSerGlnGluPheGlu 577

2901 TGGTTATGCAACCAACACT... GTTGGCGGTGTTCTCCATTTCTTTGCG 2947
|||||
577 laphPheGlnGlyLysSerLeuYrTlleasnSerGlyAsnIleProasp 593
|||||
2948 AATTCCTTCACAGGGAACACACTGACTTTGGGTGGCTTAACTT... CAG 2994
|||||
594 TyrLeuPheasp... 597
|||||
2995 TACTTTTGTGACCACCACCAAGCTTGTCATTTGTGAGGAGCATCCACTT 3044
|||||
598 PhePheGluHisLeup 603
|||||
3045 CCCAATACGAGAAATTAAGACATCACCGACAGCACTTTTTCAGTTCTGCG 3094
|||||
603 roasnCys.... AlaSerAlaIleAspPheIleLysLeuAspPheTyr 617
|||||
3095 AACAATGTTTTCACAAATCTCACGGGCCCACTATGATCAGACTATGCT 3144
|||||
618 GlyGly..... AlameIaSerTrpIleu..... 625
|||||
3145 TCTGCCCTTGAACCTATGATGATGATGAGGAGGAATTTAGCTGAAAAAGA 3194
|||||
626 LysAlaIaGlnAspT 631
|||||
3195 GCATTAATGTAAAGACTATATGATATGACAGCGAGGACATCACCAACATC 3244
|||||
631 hrcGlyGlyIleHisMetGluGlnAlaProGluThrTyr... IleProSer 646
|||||
3245 TTAGTACTGTGGCTATTGGAACTTCTCCAAAGCAGTACAGATTCCC... 3291
|||||
647 ArgAlaValSerLeuPhePheAsnTrpLysGlnIuPheArgIleuGln 663
|||||
3292 TGTCTAGA 3299
|||||
663 uvaIThrIleArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrL 680
|||||
3300 AGTCGACATGGAATGATATTGATGTTGTGAGCCAGAGATATGCTTAGATTC 3349
|||||
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
|||||
3350 TATGACACGTTTCTTCAGCTTCACGCGCCATCGAACTTCATTTAAACAC 3399
|||||
697 CysAlaGlyValAlaGlySerLeuSerIleValLeuSerThrCysLys.. 712
|||||
3400 AGCAGAGGCTTTATGAAAGGCTCCGCCACCTCTGTAGCTGTAAAGGC 3449
|||||
713 AsnIleYrSerLeuMetValGlnAlaIaSerProIleThrIleGlnAspG 729
|||||
3450 CTCGTGCACCAAGTGCTCATAGCAAGATGGAACTAGCGCACCGAC 3499
|||||
729 IuArgHisIleHrSerValThrAsnLeuLysThrLeuSerIleHis.. 744
|||||
3500 AGGAAGCTGTTCTCACCCCTGCTCCCGGAATCTTGTGAAGTCCAGG 3549
|||||
745 AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
|||||
3550 ACAATGCAATCACAAAGACCAAACTTTCTTAAT... CTGGATTAAGTT 3593
|||||
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet.. 775
|||||
3594 CCGTGTCCGTGAAGAAAGTGTCTGTGATCTGGAGGCAATTAATATTTT 3643
|||||
775 775
|||||
3644 TTTCAGTCATTCCTGAGAATTTCCAAACTTCACCAATATGAGAAATTA 3693
|||||
776 AsnGlnGluAspAlaIleLysLeuAlaGlnGlyLe 787
|||||
3694 TTGATCCAAATTTTCAAGCTGAGATGATATCTTCACCAACTAGTAAATTAAT 3743
|||||
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu..... ThrHisLys 802
|||||

```

3744 TCMAATTCGCAACCTTCATGTTTTTCATGCGAAGTGAATCTCTTT 3793
802 eTasPllEGlyGluGlyMetAspTYRllEValYsSerLeuSerSerGlu 818
|||||:|||||:|||||:|||||:
3794 CGGATTTGGGCTCTCATACATTCCTGTTTCC..... 3828
819 ProCys...AsPleuGluGluIleGluLeuValSerCysCysLeuSerAl 834
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3829 ..TGTAAAGAACTCACAGAAATTAGTTTCGGATTCATTTTTT.... 3870
834 aAsnAlaValYsllELeuAlaGlnAsnLeuHisAsnLeuValYsLeuS 851
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3871 .CAACCCCTCCCATTTGGTGTCC...AGTTGGCAAAATTTTATTTCTCTGA 3916
851 eTllELeuAsPleu...SerGluAsnTYRLeuGlnYsAsGlyAsnGlu 866
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3917 AGATTTTAAATCTTGAAAGCCACCAATTTCTCGATGAGGAACATCGAGA 3966
867 AlAlEuhHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAl 883
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3967 AATTTGGCTCATTTTATGAGTCTCTTATGTAACCTGGAAGA..... 4008
883 AlEuhMetLeuProTYRClYsAspValGlnGlySerLeuSerSerleuL 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4009 .TTGATCTTCTTACGCGGAGTGAATTTATCGACTGCGCCAAATCTGTCA 4057
900 eULYsHlELeuGluGluValProGlnLeuValYsLeuGlyLeuLYsAsn 916
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4058 TCCACAGACTGCAGACACTTCATGTCTCGAGTCCCTCTATTTTTCAG 4107
917 TrpArgLeuThrAspThr.....GluIleArgIleLeuGl 928
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4108 ACTTTGAATGATGCACCGCTGTGTGAATTTGCCAAATGACAAATCACTG 4157
928 ValAlaPheGlyLYsAsnProLeuLYsAsnPheGlnIleLeuAsnLeuA 945
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4158 AGGTTT.....CAGAACTTGAGAACCTTAAGCTT 4189
945 lAgLYsAsn..ArgValSerSerAspGlyTYRLeuAlaPheMetGlyVal 960
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4190 CAATCAATCAACAGAGATTACAGAGAGGATACGAAATTTCTTCAAGCA 4239
961 PheGluAsnLeuLYsGlnLeuValPhePheAspPheSer.....Th 974
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4240 CTGGACACATGCCAAACTTGCAGGAGTTGGACATCTCCAGGACTTTTCAC 4289
974 lLYsGluPheLeuProAspProAlaLeuValArgLYsLeuSerGlnVal 991
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4290 AGAGTGATGTCAAAGCTCATTTAGACGTGAAACATTTAAGTTGGCTCTTG 4339
991 eULSerLYsLeuThrPheLeuGlnGluAlaArgLeuValGlyTYRGlNphe- 1007
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4340 TGTTAACGCTACCAAGGCTCATTTAGACGTGAAACATTTAAGTTGGCTCTTG 4389
1008 AsPAspAsPAsPLeuSerValIle 1015
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4390 GATGCAAGTATGATTTGCAATTGCCT 4413
seq_name: /STDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX56272
seq_documentation_block:
ID AAX56272 standard; cDNA to mRNA; 5984 BP.
XX
XX AAX56272:
XX
XX 20-JUL-1999 (first entry)
XX
XX Human apoptosis inhibiting protein encoding cDNA #1.
DE
XX Human; apoptosis inhibitory protein; apoptotic disease; diagnosis;
KW
XX spinal muscular atrophy; ds.
XX
XX Homo sapiens.
OS

```

XX JP1116599-A.
PN
XX
PD 27-APR-1999.
XX
PF 14-OCT-1997; 97JP-0280831.
XX
PR 14-OCT-1997; 97JP-0280831.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI: 1999-323531/27.
DR P-PSDB: AAY09539.
XX
PT New apoptosis inhibitory protein - useful for determining mechanism
XX of various apoptotic diseases e.g. human spinal muscular atrophy
PS
XX Claim 7, Page 11-13; 16pp; Japanese.
XX
CC The present sequence encodes a human apoptosis inhibitory protein.
CC The apoptosis inhibitory protein is useful for the elucidation of
CC the mechanism of various apoptosis diseases such as human spinal
CC muscular atrophy and the diagnosis, the prevention and the treatment
CC of such diseases.
XX
XX Sequence 5984 BP; 1777 A; 1267 C; 1267 G; 1673 T; 0 other;

alignment_scores:
Quality: 653.00 Length: 1108
Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736

alignment_block:
US-09-697-089-2 x AAX56272 ..

Align seg 1/1 to: AAX56272 from: 1 to: 5984

69 PheLeuLysSerLeuLysGluTrpAsnTyrProLeuPheGlnAspLeuAs 85
1321 TTCTCCAAATATGAGTCC...TCTGCGAAGTACTCCAGACCTTCA 1367
85 nGlyIn.....SerLeuPheHisGlnThrSerGluGlyAspL 98
1368 GAGCGGTGTGAACCTTGTGAATTAAGTGAACACAGTGAAGCAATC 1417
98 euAspAsp..... 100
1418 TTGAAGATTCAATAGCAGTTGGTCTATAGTCCAGAAATGGCAGAGGT 1467
101LeuAlaGlnAspLeuLysAs 107
1468 GAAGCCCACTGTTCCAGAGCAAAAGATCTGAATGACACCTGAGAC 1517
107 PheUtyrHisThrProSerPheLeuAsnPhe..... 117
1518 ACCTTATACAGCCGACGATTTCCGCACATGCTTGTGATATCTCTT 1567
118TyrProLeuGlyGlnAspLleAspLleIlePhe 128
1568 CCGATCTGGCCAGCAGCACCTTGGGCTGTGATCTGTCTATT..... 1611
129 AsnLeuLysSerThrPheThrGlnProValLeuTrpArgLysAspGlnI 145
1612 ...GCTTCAAAACATCAGCAACCTGTG..... 1638
145 sHisHisArgValGlnGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeu 162
1639CAAGAACCTCTGTGTCTGCTGAGGCTTTGGCAACTTGA 1678
162 InSerProCysIleIleGluGlnGlySerGlyLysSerThrIleu 178
1679 ACTGTGTATGTGTGTGAGGAGGTGAAGCTGGAAGCAAGCGGTCTC 1728

179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuThr 195
1729 CTGAAGAAATATGACTTTTGTGGGCATCTGGATGCTGTCCCTGTAA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
1779 CAGGTTCAGCTGTGTCTTCTACCTCTCCCTTAGTTCCACAGACAGAG 1828
210 LysGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspLleProGlyThr 226
1829 AGGGCTGGCCAGTATCATCTGTGACCACTCTAGAGAAAGAGAGATCT 1878
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgIleArgVa 243
1879 GTTACTGAATGTGCATGAGACATTTATCCAGCATTAATAATCAGGT 1928
243 IleuPheLeuLeuAspGlyTyrAsnGlnPhe.....LysProGlnAsn 258
1929 CTATATCTTTTATGATGACTCAAGAAATATGTTCAATCCCTCAA.... 1974
258 yProGluIleGluAlaLeuIleLysGlnAsnHisArgPheLysAsnMet 274
1975GTCAAGAAACATGATTCAAAAAACCACTTATCCCGGACCTGC 2019
275 ValIleValThrThrThrGlyCysLeuArgHisIleArgGlnPheG 291
2020 CTATGATGTGTGCTGCTGCTGACAGAGCCAGGACATCCGCGATTCCT 2069
291 yAlaLeuThrAlaGlnValGlyAspMetThrGlnLysSerAlaGlnAla 308
2070 AGAGACCATTTAGATGAAAGCAATTTCCCTTATATATGCTGTGA 2119
308 euIleArgGlnValLeuIleLysGlnLeuAla.....GluGlyLeuLeu 322
2120 TATTACGAGAGCTTTTTCACATAATATGACTGCTGGAAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysTyrPro 339
2170 GTTTACTTTGAAACACCAAGTTTGGCAAGATATCAAGAAATCTCTCT 2219
339 uPheValIleThrThrCysAlaIleGlnMetGlyLysSerGlnPheHis 356
2220 CTTTGGGCGGATCTGTGCTGATTTGTCATGATCTTTTTCAGCAT 2269
356 eHisThrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
2270 CCTTGTGATGATGTGCTGTTTCAAGTCTCTATATGAAACGCTTCTCT 2319
373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspPheIleArg.. 388
2320 AGGAACAAA.....GGACACGCTGAATTTCTAAAGC 2351
389 .SerLeuAspHisCysGlyAspLeuAlaLeuGlnIlyValPheSerHisL 405
2352 AACTGTGCTCTGCTGTGTGAGCTGGCTTGAAGGTTTTCATGATGT 2401
405 ySPheAspPheGluLeuGlnAspVal.....SerSerValAsnGlnAsp 419
2402 GCTTGTGATTTAATGATGATGATTCGCAAGACGAGGGTTGATGAAG 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
2452 GAAGATCTACCATGTGCTGTGAGCAAAATTTACAGCCAGCAAGCTAAG 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnIlyThrAlaGln 453
2502 ACCATCTTACCGGTTTAAAGTCGTGCTCCAAAGAAATTTCTTGGGGGA 2551
453 rGArgLeuSerSerLeuLeuThrSerHisGlnProGluGlnValThrLys 469
2552 TGAGGCTGATGAACTCTGATTCAGATAGCAGGAACATCAAGATTTTG 2601


```
4240 CTGGACAACTGCCAACTTGACGAGTTGGACATCTCCAGCATTTTCAC 4289
974 rlysglupheleuproasprralaleuvalargylsleuserglval 991
4290 AGAGGTATCAAAAGCTCAGGCCAACACAGTCTTTGAGTCAATGCTG 4339
991 euserlysleuthrpheluglnlualargleuvalglytrpglnphe 1007
4340 TGTACGACTACCAAGGCTCATATGACGTGACATGTAAGTTGGCTCTTG 4389
1008 ASPASPaspaspserValle 1015
4390 GATGACAGATGATATGTCATTGCTT 4413

seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAA39808
seq_documentation_block:
ID AAA39808 standard; cDNA; 5984 BP.
XX
AC AAA39808;
XX
DT 22-SEP-2000 (first entry)
XX
DE Human NAIP cDNA.
XX
NAIP: apoptosis inhibiting protein; monoclonal antibody; diagnosis;
apoptosis disease onset mechanism; drugs development; prevention;
treatment; human; ss.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT CDS 292..4503
FT /tag="a
FT /product="NAIP"
FT
PN WO200024889-A1.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-JP05841.
XX
PR 26-OCT-1998; 98JP-0304550.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PA (SAKA/) SAKAI H.
XX
PI Sakai H, Ikeda J;
XX
DR WPI: 2000-350731/30.
XX
DR P-PSDB: AAY88053.
XX
PT Simple and accurate quantitation of human apoptosis inhibitory protein
(SNAIP) with monoclonal antibodies, for disease diagnosis and
development of drugs to prevent and treat apoptosis-related diseases -
Example 1: Page 31-34; 36pp; Japanese.
XX
PS
XX
CC This invention describes novel anti-NAIP monoclonal antibodies, produced
CC by hybridomas, prepared by fusing antibody-producing cells of mammals
CC immunized with an immunogen containing residues 256-586 or 841-1052 of a
CC 1403 residue amino acid sequence, fully defined in the specification, or
CC their partial sequence, with a myeloma cell line. The monoclonal
CC antibodies are for the study of apoptosis disease onset mechanism,
CC disease diagnosis and development of drugs to prevent and treat
CC apoptosis-related diseases. The quantitation method with the monoclonal
CC antibodies is simple and accurate by using a biological specimen. This
CC sequence encodes the human apoptosis inhibitory protein NAIP which is
CC described in the method of the invention.
XX
SQ Sequence 5984 BP; 1777 A; 1267 C; 1267 G; 1673 T; 0 other;
```

```
alignment_scores:
  Quality: 653.00      Length: 1108
  Ratio: 1.074        Gaps: 42
  Percent Similarity: 54.874      Percent Identity: 23.736

alignment_block:
US-09-697-089-2 x AAA39808 ..
Align seg 1/1 to: AAA39808 from: 1 to: 5984

69 pheLeuLysSerLeuLysgluTrpAsnTyrProLeuPheGlnAspLeuAs 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1321 TTCTCCAAAATATGAAGTCC...TCTGCGAAGTACTCCAGACCTTCA 1367
85 nglYgln.....SerLeuPheHtsGlnTrpSerGluGlyAspL 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1368 GAGCCGTGTGTAACCTTGTGAATTAAGTAACACAGTAAGTAAGCAATC 1417
98 euAspAsp.....
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1418 TTCAAGATTCAATAGCAGTTGGTCTATAGTCCAGAAATGGCACAGGT 1467
101 .....DeuAlaGlnAspLeuLysAs 107
1468 GAAGCCAGTGTTCAGAGGCCAAGAAATGTAATGACACGCTGAGAGC 1517
107 pleuTyrHisThrProSerPheLeuAsnPhe..... 117
1518 AGCTTATACCAGCGCCAGTTCCGCCACATGTCCTTGATATCTCTT 1567
118 .....TyrProLeuGlnGluAspLeuAspLeuPhe 128
1568 CCGATCTGGCCACGACACCACTTGGCTGTGATCTGTCTATT..... 1611
129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspLnh 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1612 ...GCTTCAAAACACATACAGCAAACTGTG..... 1638
145 shHisArgValAlGluGlnLeuThrLeuAsnngLysLeuGlnAlaLeuG 162
1639 .....CAAGAACCTGTGTGCTGCTGCTGAGGCTTTGGCACTTGA 1678
162 InseRProCysIleIleGluGlyGluSerGlyLysGlyLysSerThrLeu 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1679 ACTCTCTCATGTCTGTGAGGCGTGAAGCTGGAAGTGAAGACGCTC 1728
179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuTh 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1729 CTGAAAGAAATAGCTTTCTGTGGGCATCTGATCTGCCCTTTAA 1778
195 rlysglupheLysPheValPhePheLeuArgLeuSer.....ArgAlaGlnG 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1779 CAGGTTCCAGCTGTGTTTCTACCTCTCCCTTAGTGTCCACCAGCAGAG 1828
210 lYgLYleuPheGluThrLeuCysAspGlnLeuLeuAspLeuProGlyThr 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1829 AGGGCGTGGCCAGTATCATCTGTGACACACCTCTAGAGAAAGAGATCT 1878
227 lLeArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVa 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1879 GTTACTGGAATGATGCAATGAGGAACATTATCCAGCACTTAAGAACATCG 1928
243 lLeuPheLeuLeuAspGlyTyrAsnGlnPhe.....LysProGlnAsnG 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1929 CTATATCTTTAGATGACTACAAAGAAATATGTCATATCCCTCAA... 1974
258 yuProGluIleGlnAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1975 ....GTATAGGAAGAACTGATTCAAAAAACCACTTATCCGCGACCTCC 2019
275 ValIleValThrThrThrGluCysLeuArgGlnHisIleArgGlnPheG 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2020 CTATGATGTGCTGCTCGTACAAACAGGGCCAGGACATCCGCGATACCT 2069
```

[illegible]

```

3744 TCAAAATTCACAACTTCATGTTTCCATGCAAGTGAACCTTCTTT 3793
802 eerspiilegyluglymetasptryilevallysserleusserleu 818
3794 CGGATTTTGGGTCTCATGACTATGCTGTGTTCC..... 3828
819 Procys...AspleuglugliuleglnleuValSerCysCysleuSera 834
3829 ...TGTAGAACTCACAGAAATTAAGTTTCGATTCATTTT..... 3870
834 aasnAlaVallyslleuAlaInleuH1asnleuH1asnleuVallysl 851
3871 .CAAGCCGTCCTCATTTGTTCC...AGTTGCCAAATTTATTTCTCTGA 3916
851 erlleuaspLeu...SerGluasnTyrlleuGlulysaspGlyAsnGlu 866
3917 AGATATTAAATCTTGAAGCCACCAATTTCTGTATGAGGAACATCGAA 3966
867 AlaLeuH1sgluLeuLeaspArgmetAsnValleuglnleuThrAl 883
3967 AAATTTGCCACATTTTAGTCTCTTAGAACCTGAGAA..... 4008
883 aleuMetleuProTrrpGlyCysaspValglnGlySerleuSerleu 900
4009 .TTGATCCCTTCCTAGCGGATGGAATTTATCGAGTGCCAAACTGATCA 4057
900 eulysH1sleuglnGluValPrrglnleuVallyslleuGlyLeuYasn 916
4058 TCCAGCGTGTCCAGACGCTCATTTGTCGAGTCTCTCATTTTTCAG 4107
917 TrpArgleuThrAspThr.....GluIleArglleuGln 928
4108 ACCTTGAATATGACACCGTGTGGAATTCGCAAACTTACCAATCAGTGG 4157
928 yAlaPhePheGlyLysAsnProleuYasnPheGlnleuAsnleuA 945
4158 AGGTTTC.....CAGAAACTGAGAACCTTAAGCTTT 4189
945 laGlyAsn...ArgValSerSerAspGlyTrpLeuAlaPheMetGlyAl 960
4190 CAATCAATCAACAAAGATTACAGAGGAAGATTAAGAAATTTCTTCAAGCA 4239
961 PheGluAsnleuYsglnleuValPhePheAspPheSer.....Th 974
4240 CTGGACACACATGCCAAACTTGCAGAGATTGACATCTCCAGCATTTTCA 4289
974 rLysGluPheLeuProaspProAlaLeuValArgLysLeuSerGlnVal 991
4290 AGAGTGTATCAAGCTCAGGCCACACAGCAAGTCTTTGAGTCAATGTG 4339
991 euserLysLeuThrPheLeuGlnGluAlaArgLeuValGlyTrpGlnPhe 1007
4340 TGTATGACATCAACAGCGCTATTAGACTGAACATGTTAAAGTGGCTCTTG 4389
1008 AspAspAspAspLeuSerValIle 1015
4390 GATGACGATCATATTCATTCCTT 4413
seq_name: /SID52/gc9data/geneseq/geneseqn/NA1997.DAT:AAAT71265
seq_documentation_block:
ID AAAT71265 standard; cDNA: 6124 BP.
XX
XX AAAT71265;
XX
XX 06-OCT-1997 (first entry)
XX
XX Neuronal apoptosis inhibitor protein (NAIP) cDNA.
XX
XX Neuronal apoptosis inhibitor protein; NAIP; diagnosis:
XX gene therapy; cancer; AIDS; amyotrophic lateral sclerosis;
XX spinal muscular atrophy; ds.

```

```

XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 292..4503
FT CDS /*tag= a
XX
XX W09726331-A2.
XX
XX 24-JUL-1997.
XX
XX 17-JAN-1997; 97WO-IB00142.
XX
XX 19-JAN-1996; 96GB-0001108.
XX
XX (UNOT-) UNIV OTTAWA.
XX
XX Korneluk RG, Mackenzie AE, Robertson G, Roy N, Tamai K;
XX WPI: 1997-385335/35.
XX P-PSDB; AAM20032.
XX
XX New neuronal inhibitor of apoptosis - useful for diagnosing and
XX treating, e.g. cancer, AIDS or amyotrophic lateral sclerosis
XX
XX Claim 28; Fig 6A-I; 102pp; English.
XX
XX A cDNA clone (AAAT71265) codes for novel human neuronal apoptosis
XX inhibitor protein (NAIP) (AAM20032). It was isolated from a human
XX foetal spinal cord cDNA library by probing with the genomic insert
XX in cosmid 25086, containing a CAT locus, and completion of gaps
XX in the sequence by further library screenings. The NAIP gene, on
XX chromosome 5q13.1, is about 56 kb long and has 18 exons (see also
XX AAAT71266). NAIP is a negative regulator of apoptosis, partic.
XX neuronal apoptosis and, when deficient or absent, contributes to
XX neurodegenerative phenotypes such as spinal muscular atrophy (SMA)
XX and amyotrophic lateral sclerosis. NAIP nucleic acids are useful
XX for protein expression in host-vector systems, as probes/primers,
XX and in gene therapy to inhibit apoptosis for treatment of AIDS,
XX neurodegenerative disease, myelodysplastic syndromes or ischaemic
XX injury, or (antisense) to increase apoptosis.
XX
XX Sequence 6124 BP; 1811 A; 1297 C; 1286 G; 1730 T; 0 other;
SQ
alignment_scores:
Quality: 653.00 Length: 1108
Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736
alignment_block:
US-09-697-089-2 x AAAT71265 ..
Align seg 1/1 to: AAAT71265 from: 1 to: 6124
69 PheLeuLysSerLeuYsglnTrpAsnTyrrProLeuPheGlnAspLeuAs 85
1321 TTTCTCCAAATATGAAAGTCC...TCTCGGAAAGTACTCCACACCTTCA 1367
85 nGlyGln.....SerLeuPheH1sglnThrSerGlnGlyAspL 98
1368 GAGCCGTGTGGAACCTTGTGAATTACTGGAACACCAAGGAAGCAATC 1417
98 euAspAsp..... 100
1418 TTGAGATTCAATATGACGTGTGCTATAGTCCAGAAATGGCACAGGCT 1467
101 .....LeuAlaGlnAspLeuYsas 107
1468 GAAGCCAGTGTTCACAGAGCAAGAAATCTGAATGACGAGCTGAGAC 1517
107 PleuTyrrH1sThrProSerPheLeuAsnPhe..... 117

```

1518 AGCTTATACACAGCCAGTTTCGCCACATGCTTGTGATATCTCTT 1567
118TyrProLeuGluGluAspIleAspIleLeuPhe 128
1568 CGGATCTGCCACGACGACCTTGTGGGCTGTGATCTGTCTATTT 1611
129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnI 145
1612 ...GCTCAAAAACATCAGCAAACTGTG..... 1638
145 ShisHisArgValGluGlnLeuThrLeuAsnGlyLeuGlnAlaLeuG 162
1639CAGAACTCTGTGCTGTGCTGAGAGTCTTTGGCAACTTGA 1678
162 InseProCysIleIleGluGluGluSerGlyLysGlyLysSerThrLeu 178
1679 ACTGTCTGATGTGTGTGGAGGTGAGCTGGAAGTGAAGAGCGTCTCTC 1728
179 LeuGlnArgIleAlaMetLeuTrpLysSerGlyLysCysLysAlaLeuTh 195
1729 CTGAAGAAATAGCTTTCTGTGGGCATCTGATGCTGCTCCCTGTAA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGlnG 210
1779 CAGGTTCACAGCTGTTTCTACTCTCTCCCTTAGTTCACACAGACAGACG 1828
210 lGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226
1829 AGGGCTGGCCAGTATCATCTGTGACCACTCCAGAAAGAGAGATCT 1878
227 lLeuArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArg 243
1879 GTTACTGAATGTGCATGAGGAACATTTATCCAGATTAAAGAACATCAGCT 1928
243 lLeuPheLeuLeuAspGlyTyrAsnGlnPhe.....LysProGlnAsnG 258
1929 CTATTCTCTTATGATGACTACAAAGAAATATGTCATCTCTCA.... 1974
258 yspProGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
1975GTCATAGGAACACTGATTCAAAAAACCTTATCCCGACCTGCG 2019
275 ValIleValIleThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
2020 CTATTGATGTGCTCGTACAAACAGGCCAGGACATCCGCCGATACCT 2069
291 yAlaLeuThrAlaGluValAlaGlyAspMetThrGluAspSerAlaGlnAla 308
2070 AGAGACCATTCAGAGATCAAGCATTTCCCTTATATATACTGTCTGA 2119
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
2120 TATTCAGGAAGCTCTTTACATATATGACTGCTGTGCAAAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProLe 339
2170 GTTATCTTGGAAAGAACCAAGTTTGCAGAGATACAGAAACCTCTCT 2219
339 uPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
2220 CTTTGTGGCGGAGATCTGTGCTGATGTTTCAGTATCTCTTTGACCAT 2269
356 eHisThrGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
2270 CCTTGTGATGATGTGGCTGTTTCAAGTCTTATGTAAGCGCTTCTCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaAlaAspAspPheIleArg. 388
2320 AGGAACAAA.....GCGACAGCTGAATTTCTCAAGC 2351
389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGlnGlyValPheSerHis 405
2352 AACGTGTCTCTGTGTGAGCTGCTTGAAGGCTTTTTCATGTT 2401

405 yspPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
2402 GCTTGAAGTTTAAATGATGATGATCTCGCAGAAAGCGGGGTGATGAACAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
2452 GAAGATCTAACCTGTGCTTATGATGAGCAAAATTTACAGCCAGAGACTAAG 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnGluThrArgAlaGly 453
2502 ACCATTTACCGGTTTAACTCTGCTGCTCCAGAAATTTCTTGCGGGA 2551
453 rArgLeuSerSerLeuLeuThrSerHisGluProGluValThrLys 469
2552 TGAGCGCTATGAACTCTCGATTCAGATGAGCAGAACATCAAGATTG 2601
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
2602 GGACTGTATCATTTGAACAAATCAACTACCCATGATGACTGTGAAGCG 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
2652 CTACACAAATTTTGAACATATGTC.....TCCAGCTCCCTTCACAA 2695
503 rGla.....ValMetLysHisLeuAlaAlaValTyrGlnHisGly 516
2696 AAGCAGGCCCAAAATGTGTCTCATTTGCTCCATTTAGTGATGATAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnGluSe 533
2743AAAGATC 2750
533 rLeuGlnSerValLysAsn..... 539
2751 ATTGAGAAATATATCTGAAATGATGACTACTTAAGCAGCACAGCAAA 2800
540 ..ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551
2801 TTTCACCTCAGATGCAGTTACTTAGGGGATGTGCAAAATTTGCCACAA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2851 GCTTACTTTTCAATGTTTCAAGAACATTTACTGTTCTTGCCCTGAAGAC 2900
562TyrGlnLysSerThrSerLysSerAlaLeuSerGlnLysPheGln 577
2901 TGCTTATCAAGCAACT...GTTCGTGCGTGTCTGCTCATTTGTTTGC 2947
577 lArgPheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsp 593
2948 AATTCCTTCAAGGGAGAACACTGACTTGGGTGGCCTTAACTTA...CAG 2994
594 TyrLeuPheAsp..... 597
2995 TACTTTTTCAGACCCAGAAAGCTTGTCATTTGTTGAGGACATCCACTT 3044
598PhePheGlnHisLeuP 603
3045 CCCAATACAGGAATAAGACATCACCCAGAGACATTTTTCAGTTCTGG 3094
603 rAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
3095 AAACATGTTTTCAGAAATCAGACAGTGCCAACTATAGATCAGGACTATGCT 3144
618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
3145 TCTGCTTTGAACCTATGATGAATGAGGAGCAAAATTTAGCTGAAGAAAG 3194
626LysAlaIleGlnAspT 631
3195 GGATATGTAAAGAGCTATATGATATGACAGGACAGGACATCACAGAGC 3244


```

3299 GGATAATGTAAGACTATATGATATGCAGCGAGGGCATCCACGACC 3348
631 hrglyglyIleHisMetGluGluIuaIaProGluThrYr...IleProSer 646
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3349 TTAGTACTGGCTATTGGAACTTCTCCAAAGCAGTACAGATTCCC... 3395
647 ArgAlaValSerLeuPhePheAsnTrpLysGlnIupheArgThrLeuI 663
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3396 .....TGTCTAGA 3403
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrYrL 680
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3404 AGTCAGATGTGATGATATGATGTGTAGCGCAGAGATGTGCTGAGATT 3453
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3454 TAAATGACAGTTTCTCAGCTTCACAGCGCATCGAACTCCATTAAACAC 3503
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3504 AGCAGAGGCTTTATAGAAAGCATCCGCCACGCTGTGAGCTGTAAAGGC 3553
713 AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAspG 729
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3554 CTCTGTACCAAGTGTCTCCATAGCAAGTTGGAACTCAGCGCAGCCGAC 3603
729 IuaArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3604 AGAAAGTGTCTTCACCCCTGCCCTCCGGAATCTCTTGAAGTCTCAGGG 3653
745 .....AspLeuGlnAsnGlnArgLeuProGlyLysLeuThrAspSerLe 759
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3654 ACAATCCAGTCCACAAGACCAAAATCTTCTCTAAT.....CTGATTAAGTT 3697
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnLysMet.. 775
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3698 CCTGTGCTGTAAGAACTGTGTGTGATCTGAGGCGCAATATAATGTTT 3747
775 ..... 775
3748 TTTCAGTCATTCCTGAGAATTTCCAAACTCCACCATATGAGAAATTA 3797
776 .....AsnGluGluAspAlaIleLysLeuAlaGluGlyLe 787
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3798 TTGATCCAAATTTGAGCTGAGTATGATCTTCCAAACTAGAAATTAAT 3847
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeus 802
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3848 TCATAATTTCCAAACCTTCATGTTCATCTGTAAGTAACTTCTTTT 3897
802 eTrAspIleGlyGluGlyMetAspTrpIleValLysSerLeuSerSerGlu 818
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3898 CGGATTTTGGGTCTCATGACTGATGCTTGTTC..... 3932
819 ProCys...AspLeuGluGluIleGlnLeuValSerCysCysLeuSerAl 834
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3933 ...TGTAAGAAACACAGAAATTAAGTTTCGGATTCATTTT... 3974
834 aAsnAlaValLysIleLeuAlaGlnAsnLeuHisIleLeuValLysLeus 851
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3975 .CAAGCCGTCCTCATTTGTGCC...AGTTGCCAAATTTATTTCTCTGA 4020
851 eTrIleLeuAspLeu...SerGluAsnTrpLeuGluLysAspGlyAsnGlu 866
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4021 AGATAATTAAATCTTGAAGCCAGCAATTTCTGTAGAGGAACAATCAGAA 4070
867 AlaLeuHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAl 883
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4071 AAATTTCCCTACATTTAGTTCTCTTAACTGTAACCTCGAAGAA..... 4112
889 aLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuL 900
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4113 .TTGATCTTCTTACTGTGGGATGGAATTTATGAGTGGCCAAACTGATCA 4161

```

```

900 euLysHisLeuGluGluIuaIaProGlnLeuValLysLeuGlyLeuLysAsn 916
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4162 TCCAGCACTGTCCAGACCTTCATGTCTCCGAGTCCTCTCATTTTTCAG 4211
917 TrpArgLeuThrAspTrpHisGluIleArgIleLeuGlyAlaPhePheGly 933
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4212 ..ACTTTGAATGATGACACGCTGTGGAAATTTGCCAAATGACAAATCA 4258
933 sAsnProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsn...ArgV 949
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4259 TGGAGTTTCCAGAAACTTGAGAACTTAAGCTTTCAATCAATCAAGAA 4308
949 aLysSerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLys 965
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4309 TTACAGAGGAAGATACAGAAATTTCTTCAAGCACTGACAAACATGGCA 4358
966 GlnLeuValPhePheAspPheSer.....ThrLysGluPheLeuPr 979
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4359 AACTTGCAGAGTGTGACATCTCCAGGCAATTCACAGAGTGTACAAAGC 4408
979 oAspProAlaLeuValArgLysLeuSerGlnValLeuSerLysLeuThrP 996
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4409 TCAGGCCACAACAGTCAAGTCTTTGAGTCAATGTGTACGACTACCAA 4458
996 heLeuGlnGluAlaArgLeuValGlyTrpGlnPheAspAspAspLeu 1012
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4459 GGCATATAGACTGAAACATGTTAAGTTGCTCTGTGATGACAGATGATATT 4508
1013 SerValIle 1015
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4509 GCATTTGCTT 4517

```

seq_name: /STDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.AAX58001

seq_documentation_block:

ID AAX58001 standard; cDNA to mRNA; 5366 BP.

AC AAX58001;

DT 20-JUL-1999 (first entry)

DE Gonadotropic hormone coding sequence.

KW Gonadotropic hormone; excessive ovulation animal; transgenic animal;

KM totipotent cell; somatic cell chromosome; ds.

OS Homo sapiens.

PN JP11113444-A.

PD 27-APR-1999.

PF 14-OCT-1997; 97JP-0280830.

PR 14-OCT-1997; 97JP-0280830.

PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.

PA (SAKA/) SAKAI H.

DR WPI; 1999-320709/27.

DR P-PADB; AAY14080.

PT An excessive ovulation animal - useful for improving the

productivity of animals

PS Claim 1; Page 9-11; 18pp; Japanese.

CC This sequence encodes a gonadotropic hormone.

CC The invention relates to an excessive ovulation animal, which is a

CC transgenic animal with a totipotent cell containing a DNA fragment

CC containing a promoter sequence and a gonadotropic hormone coding

CC sequence. The DNA fragment is in the somatic cell chromosome. The

CC excessive ovulation animal is useful for improving the productivity of
CC animals. The method can improve the productivity of a useful animal.
XX
SQ Sequence 5366 BP: 1483 A; 1173 C; 1222 G; 1488 T; 0 other;

Alignment_scores:

Quality: 593.50 Length: 1094
Ratio: 1.023 Gaps: 42
Percent Similarity: 53.016 Percent Identity: 23.675

Alignment_block:

US-09-697-089-2 x AAX58001 ..

Align seg 1/1 to: AAX58001 from: 1 to: 5366

```
69 PheLeuLysSerLeuLysGluThrPasnTrpProLeuPheGlnAspLeuAs 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1321 TTTCTCCAAATATGAGATCC...TCTCGGAGAGTACTCCAGACCTTCA 1367
85 nGlyGln.....SerLeuPheHisGlnPheSerGluGlyAspL 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1368 GAGCGGTGGTGAATTTGTGAATTAAGTGAACACACAGTGAAGCAATC 1417
98 euAspAsp..... 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1418 TTGAAGATTTCATAGCAGTTGGTCTATAGTCCAGAAATGSCACAGGT 1467
101 .....LeuAlaGlnAspLeuLysAs 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1468 GAAGCCAGTGGTTTCAAGAGGCAAGAAATCTGAATGAGCAGCTGAGAGC 1517
107 PheUtyrHisThrProSerPheLeuAsnPhe..... 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1518 ACCTTATACAGCGCATTTTCCGCACATGTCCTTGTCTGATATCTCTT 1567
118 .....TyrProLeuGluLysAspLysPheAspLysPhe 128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1568 CCGATCTGGCCAGCGGACCTGCGGTGGGTGTGATCTGTCTATTT..... 1611
129 AsnLeuLysSerThrPheThrGluProValLeuTyrPargLysAspGln 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1612 ...GCTTCAAAACACATCAGCAAACTGTG..... 1638
145 sHisHisArgValGluGlnLeuThrLeuAsnGlyLeuGlnAlaLeuG 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1639 .....CAGAAACCTCTGTGTGCTGCTGAGGCTTTGGCAACTTGA 1678
162 InSerProCysIleIleGluGluLysSerGlyLysGlyLysSerThrLeu 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1679 ACTCTGTATCTGTGTGGAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1728
179 LeuGlnAlaGluIleAlaMetLeuThrPheLysGlyLysCysLysAlaLeu 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1739 CTGAAGAAATAGCTTCTGTGTGTCATCTGATCTGATCTGATCTGATA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1779 CAGGTTCAGCTGTGTGTCTTCTACCTCTCCCTTATGTACACAGACAGAG 1828
210 LysGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspLysPheGluThr 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1829 AGGGCTGGCCAGTATCATCTGTGACAGCTCTGAGAAAGAAAGAGATCT 1878
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVal 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1879 GTTACTGAATCTGCATGAGCAATTTATCCAGCAGTTAAAGAACAGAGT 1928
243 IleuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1939 CTTATCTCTTTAGATGACTACAAAGAAATATGTTCATCTCCCTCA... 1974
258 yspProGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
```

```
1975 .....GTATAGGAAACATGATTCAAAAAACCACTTATCCCGAGCTCC 2019
275 ValIleValThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2020 CTATTGATTTGCTGCTCCGTATCAAAACAGGCCAGGACATCCCGGATCT 2069
291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2070 AAGACCACTTATAGATCAAAAGCATTTCCCTTTATATATAGTCTGTGA 2119
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeu 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2120 TATTACGGAAGCTCTTTTCATATATATAGCTGCTGTGGAAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrPro 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2170 GTTTACTTTTGGAAAGAACCAAGTTTGCAGAGATACAAAGAAATCTCT 2219
339 uPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2220 CTTTGTGGCGGAGTGTGTCTCATTTGTTCAATGCTTTTGAACCCAT 2269
356 eHisIleThrGlnThrThrLeuPheHisThrPheTyrAspLeuIleGln 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2270 CCTTGTGATGATGTGGGTGTTTCAAGTCCCTATATGAGACGCTTCTCT 2319
373 LysAsnLysHisLysHisLysGluValAlaAlaSerAspPheIleArg 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2320 AGCAACAA.....GCCACAGCTGAATTTCTCAAAAC 2351
389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGluValPheSerHis 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2352 AACTGTCTCTCTGCTGAGCTGCTGAAAGGCTTTTTCATGTT 2401
405 yspPheAspPheGluLeuGlnAspVal.....SerSerValAsnGlu 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2402 GCTTGTGATTTATGATGATGATGATGATGATGATGATGATGATGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2452 GAAGATCTAACCATGCTGCTGATGAGCAAAATTTACAGCCCAAGACTA 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaG 453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2502 ACCATTCTACCGGTTTATAGTCTGCTCCCTCCAGAGATTTCTTGGGG 2551
453 rArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2552 TTAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2601
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSer 486
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2602 GCACGTATCATTTGAAACAAATCAACTCAACCATGATGATGATGATG 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2653 CTACAAACATTTTGAACATATGTC.....TCCAGCTCCCTTAAACAA 2695
503 rGluAla.....ValMetLysHisLeuAlaAlaValTyrGlnHisGly 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2696 AAGCAGGGCCCAAAATTTGTGTCTCATTTGTCTCATTTAGTGAATAC 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTyrPargGlnGlu 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2743 .....AAAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2751 ATTGAGAAATATATCTGAATATGATGATGATGATGATGATGATGATG 2800
540 ..ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

2801	TTTCACTGCAGATGACGTACTTAGGGATTGTGGCAAAATTTCGCCACA	2850
552	AsnSerPheValGluCysGlyIleHisLeu.....	561
	:::	
2851	GCTTAATTTCATGGATTGCAGAACATTAATGTTTGCCCTGAAC	2900
562	...TYRGInGlnSerThrSerLysSerAlaLeuSerGlnGluPheGlu	577
	::: :::	
2901	TGCTTATCAAAGCAACACAC...GTGGCGCGGTTCTCATTTGTTTGC	2947
577	LapPheGlnGlnLysSerLeuTyrlIeasSerGlyYasnIleProasp	593
	::: :::	
2948	AATTCCTTAAGAGGAGACACTGACTTTGGTGCCCTTAACTTA...CAG	2994
594	TyrLeuPheAsp.....	597
	:::	
2995	TACTTTTTCAGCACCCAGCAAGAAGCTTGATCTGTGGAGAGATCCAATT	3044
598PheGlnHisLeuP	603
	:::	
3045	CCCATAACGAGAAATAGACATCACCGACACATTTTTCAGTTCTGG	3094
603	roaspYs.....AlaSerAlaLeuAspPheIleLysLeuAspPheYr	617
	::: :::	
3095	AAACATGTTTTGACAAATACACAGGTGCCAATATATGATCAGACTATGCT	3144
618	GlyGly.....AlaMetAlaSerTrpGlu.....	625
	:::	
3145	TCGTGCTTTGAACCTATGAATGAATGGAGCGCAATTTAGCTGAAGAAA	3194
626LysAlaIleGluAspT	631
	::: :::	
3195	GGATAATGTAAAGCTATATGATATGCAGCGCAGGCATCACCGACC	3244
631	hrcIglyIleHisMetGlnGluAlaProGluThrTry...IleProser	646
	::: ::: :::	
3245	TTTGTACTGGCTATTTGGAACTTTCTCCAAAGCACTAACAGATTGCC...	3291
647	ArgAlaValSerLeuPhePheAsnTrpLysGlnGluPheArgThrLeuG1	663
3292TGCCTAGA	3299
663	vUaIThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleHrTryL	680
	::: ::: :::	
3300	AGTCAGTGAATGATATTGATGTTGTAGCGCCAGATATGCTTGAGATTG	3349
680	eudLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg	696
	::: ::: :::	
3350	TAATGCAAGTTTTCACGCTTACAGCGCATCGAATCTCATTTAAACAC	3399
697	CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys..	712
	::: ::: :::	
3400	AGCAGAGGCTTATAGAAGATCCGCCACAGCTGTGAGCTGTCTAAGCC	3449
713	.AsnIleYrSerLeuMetValGluAlaSerProLeuThrIleGluaspG	729
	::: :::	
3450	CTCTGTACCAAGTGTCTCATAGCAAGATGGAACTCAGCCAGCCGAC	3499
729	IuArgHisIleHrSerValIThrAsnLysThyLeuSerIleHis...	744
	::: ::: :::	
3500	AGGAATGCTTCTCACCCGCTTCCGTGAATCTTGTGAAGTCTCAGG	3549
745AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe	759
	::: ::: :::	
3550	ACATTCACATCAACAGACCAATCTTCTTAAT.....CTGATTAAGTT	3593
759	ucgIAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet.	775
	::: ::: :::	
3594	CCGTGCTCTGAAGAACTGTCTGTGATCTGGAGGCAATTAATATGTTT	3643
775	775
6644	TTTTCACTACTTCTGAGAANTTCCAAATTCACCATATGAGAAATTA	3693

```

776 .....AsnGluGuaAspAlaIleUysLeuAlaGluGlyLe 787
              ||| ||| ||||| : : : : :
3694 TTGATCCAAATTTACAGTGAATGATGCTCCAAACGTAAATTAAT 3743
uUysAsnLeuUysUysMetCysLeuPheHisLeu.....ThrHisLeuS 802
||||| : : : : : : : : : : : : : : : : : : : : :
3744 TCAAAATTCTCCAAACCTTCATGCTTTTCCATGTGAAGTGAATCTTTT 3793
802 eArpIleGluGluGlyMetAspTyrIleValUysSerLeuSerSerGlu 818
||||| : : : : : ||| : : : : : : : : : : : : : : :
3794 CGGATTTGGTCTCTCATGACATACCTGTTTCC..... 3828
819 ProCys...AspLeuGluGluIleGlnLeuValSerCysCysLeuSerAl 834
      ||| ||| ||||| : : : : : : : : : : : : : : :
3829 ...TGTAAGAACTCACAGAAATTAAGTTTGGATTCATTTT... 3870
834 aAsnAlaValUysIleLeuAlaGlnAsnLeuHisAsnLeuValUysLeuS 851
      : : : : : : : : : : : : : : : : : : : : :
3871 .CAAGCGGTCCTCTGTTGTTCC..AGTTGGCAATTTTATTTCTCTGA 3916
851 eTlleUysAspLeu...SerGlnAsnTyrLeuGlnUysAspGlyAsnGlu 866
      : : : : : : : : : : : : : : : : : : : : :
3917 AGATATTAAATCTTGAAGCCACAGCAATTTCTCGATGAGAAACATCAGA 3966
867 AlaLeuHisGluLeuLeuAspArgMetAsnValLeuGlnGlnLeuThrAl 883
3967 AAATTTGGCTACATTTTGAAGTTCTCTGATCACTTGGAAGA..... 4008
883 aLeuMetLeuProTrrGlyCysAspValGlnGlySerLeuSerSerLeuL 900
      ||||| : : : : : ||| : : : : : : : : : : : : : :
4009 .TTGATCTCTCTCTAGGGGATGGAAATTATGAGAGGCCAAGATGATCA 4057
900 eUlysHisLeuGlnGluValProGlnLeuValUysLeuGlyLeuUysAsn 916
      : : : : : : : : : : : ||| ||| : : : : : : : : :
4058 TCAGAGAGGTCCAGCAGCTTCATGTCTCCGAGCTCTCATATTTTCAAG 4107
917 TrpArgLeuThrAspThrGlnIleArgIleLeuGly..... 928
4108 ...ACTTTGATGATGACACAGCGGTGGAAATTTGGTAGCTAGTGTTC 4154
929 ..... 930
4155 GCTTGCATGGAAGCCAGTGTATAGCCAAAGCTTTCTGTCGAACATGCTC 4204
930 hePheGlyUysAsnProLeuUysAsnPheGln.....Gln 941
      : : : : : ||| : : : : : ||||| : : : : :
4205 ATGTAAACATTTGCCCCCTAGAAATTTTCAACCCGCTTCTCATTTTCA 4254
942 LeuAsnLeuAlaGlyAsnArgValSerSerAspGlyTrrPheAlaPheMe 958
      ||| : : : : : : : : : ||| : : : : : |||||
4255 CTATCATACTGTTCCCTTAGTGTCTCTCTGTGGATTTAGGGCATTCG 4304
958 tGlyValPheGluAsnLeuUysGlnLeuValPheAspPheSerThr. 974
      : : : : : : : : : ||| ||| ||| : : : : : : : : :
4305 GTCAGATTTGGAAGTACAAAGAGGCTCCCATTTGTGGATATATACAAGCCC 4354
975 ....LysGluPheLeuProAspProAlaLeu 983
4355 TCAAATCTGCGTTCTGCAACCTGATGTTTTA 4386
seq_name: /SID82/gcdata/geneseq/geneseqn/NAI99.DAT:AA56273
seq_documentation_block:
ID   AA56273 standard; cDNA to mRNA; 5366 BP.
XX   AA56273:
AC
XX
XX
DT   20-JUL-1999 (first entry)
XX
DE   Human apoptosis inhibiting protein encoding cDNA #2.
XX
KN   Human; apoptosis inhibitory protein; apoptotic disease; diagnosis;

```



```

2552 TGAGCGTATTGAACTCTCGATTACAGATGAGCAAGCAATCAAGATTG 2601
      ||||| :||| ||| :|||
470 GlyslnGlyThrLeuGlnLysMetValSerIleSerAspIleThrSer 486
      ||| :||| :||| :|||
2602 GGACTGATATCTTTGAAACAATCACTACCCATGATGACTGATGAGCGC 2651
      ||| :||| :||| :|||
486 rTyrsSerLeuLeuArgTyrrhCysGlySerSerValGluLathra 503
      :||| :||| :||| :|||
2652 CTACACAATTTTGTGAACATATGTC.....TCCAGCCTCCCTCAACAA 2695
      :||| :||| :||| :|||
503 rGluA.....ValMetLysHisLeuAlaValTyrrhHisGly 516
      :||| :||| :||| :|||
2696 AAGCAGGGCCCAAAATGTGTCTCATTTGCTCCATTAGTGGAATAC... 2742
      :||| :||| :||| :|||
517 CysLeuLeuGlyLeuSerIleAlaLysrProLeuTrpArgGlnLys 533
      :||| :||| :||| :|||
2743 .....AAGAGTC 2750
      :||| :||| :||| :|||
533 rLeuGlnSerValLysAsn..... 539
      :||| :||| :||| :|||
2751 ATTGAGAAATATATCTGAAATGATGACTACTTAAGCACCAGCAAA 2800
      :||| :||| :||| :|||
540 ..ThrThrGlnGlnLileuLysAlaIle.....AsnIle 551
      :||| :||| :||| :|||
2801 TTTCACCTGCAGATCGAGTTACTTAGGGGATTGTGCAGAAATTTGCCAA 2850
      :||| :||| :||| :|||
552 AsnSerPheValGlyCysGlyIleHisLeu..... 561
      :||| :||| :||| :|||
2851 GCTTACTTTCAATGTTTCAGACATTTACTGTCTTGCCCTGAAAC 2900
      :||| :||| :||| :|||
562 ....TyrrhGlnSerThrSerLysSerAlaLeuSerGlnLupheGlu 577
      :||| :||| :||| :|||
2901 TGCTTATCAAGCAACACT...GTTCGTCGTGTCTCCATTGTGTTGC 2947
      :||| :||| :||| :|||
577 LaphPheGlnGlyLysSerLeuTyrrhLeuAsnSerGlyAsnIleProasp 593
      :||| :||| :||| :|||
2948 AATTCCTTCAAGGAGACACTGACTTGGGTGGCTTAACCTTA...CAG 2994
      :||| :||| :||| :|||
594 TyrrhLeuPheAsp..... 597
      :||| :||| :||| :|||
2995 TACTTTTTCGACCCAGAAAGCTTGTCAATTGTGAGGACATCCACTT 3044
      :||| :||| :||| :|||
598 .....PhePheGlnHisLeuP 603
      :||| :||| :||| :|||
3045 CCCAATACAGAAATAGACATCACCCAGACACATTTTTCAGTTCTCG 3094
      :||| :||| :||| :|||
603 rAsnGlyCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
      :||| :||| :||| :|||
3095 AATCATGTTTTCACAATTCACAGGTGCCACATATAGATCAGGACTATGCT 3144
      :||| :||| :||| :|||
618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
      :||| :||| :||| :|||
3145 TCTGCTTTTGAACCTATGAATGATGGAGGAAATTTAGCTGAAGAAAGA 3194
      :||| :||| :||| :|||
626 .....LysAlaAlaLeuAspTr 631
      :||| :||| :||| :|||
3195 GGAAATATGTAAGAGCTATATGATATGACCGCAGGGGACATCACCAACC 3244
      :||| :||| :||| :|||
631 hrGlyGlyIleHisMetGlnGluLapProGlnuThrTyrr..IleProSer 646
      :||| :||| :||| :|||
3245 TTACTGACTGGCTATTTGAAACTTTCTCCAAAGCAGTACAGATTC... 3291
      :||| :||| :||| :|||
647 ArgAlaValSerLeuPheAsnTrpLysGlnLupheArgThrLeuGlu 663
      :||| :||| :||| :|||
3292 .....TGTCTAGA 3299
      :||| :||| :||| :|||
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrrL 680
      :||| :||| :||| :|||
3300 AGTCGATGTAATGATATGATTTGTAGGCCAGGATATGCTTGAGATTC 3349
      :||| :||| :||| :|||
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
      :||| :||| :||| :|||

```

```

3350 TAATGACAGTTTCTCAGCTTCACAGCGCATGCAACTCCATTAAACCAC 3399
      :||| :||| :||| :|||
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
      :||| :||| :||| :|||
3400 AGCAGAGGCTTTATAGAAAGATCCGCCACACTCTTGAGCTGTCTAAGC 3449
      :||| :||| :||| :|||
713 .AsnIleTyrsSerLeuMetValGluAlaSerProLeuThrIleGluAsp 729
      :||| :||| :||| :|||
3450 CTCTGTCCACCAAGTGTCTCATACCAAGTGAACCTGCACTGACGACCGAAC 3499
      :||| :||| :||| :|||
729 LuArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
      :||| :||| :||| :|||
3500 AGGAATGCTCTTCACCTGCTCCCTGGAATCTTGAAGTCTCAGCG 3549
      :||| :||| :||| :|||
745 .....AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
      :||| :||| :||| :|||
3550 ACATTCACGTCCACAAAGCAATCTTCTCTAT.....CTGGATTAAGTT 3593
      :||| :||| :||| :|||
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet. 775
      :||| :||| :||| :|||
3594 CCTGTGCTGTGAAGAAACTGTCTGTGATCTGAGGGCAATATTAATGTTT 3643
      :||| :||| :||| :|||
775 ..... 775
      :||| :||| :||| :|||
3644 TTTCAGTCATTTCTGAGCAATTTCCAAACTTCACACCATATGAGCAAAATTA 3693
      :||| :||| :||| :|||
776 .....AsnGlnGluAspAlaIleLysLeuAlaGlyLe 787
      :||| :||| :||| :|||
3694 TTGATTCCAATTTTCAGCTGATGATGCTTCCAAAGTAAATATATAT 3743
      :||| :||| :||| :|||
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeus 802
      :||| :||| :||| :|||
3744 TCAAAATTTCTCCAAACCTTCATGTTTCCAACTGAAAGTAACTTCCTTTT 3793
      :||| :||| :||| :|||
802 erAspIleGlyGlyLysMetAspTyrrhLeuValLysSerLeuSerGlu 818
      :||| :||| :||| :|||
3794 CGGATTTTGGTCTCTCATGACTATGCTTGTTC... 3828
      :||| :||| :||| :|||
819 ProCys...AspLeuGlnGluIleGlnLeuValSerCysCysLeuSerAl 834
      :||| :||| :||| :|||
3829 ...TGTAAGAACTCACCAAAATTAAGTTTCGGATCATTTT... 3870
      :||| :||| :||| :|||
834 AsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeus 851
      :||| :||| :||| :|||
3871 .CAAGCCCTCCCATTTGTGCC...AGTTGCCAAATTTTCTCTCGA 3916
      :||| :||| :||| :|||
851 erIleLeuAspLeu...SerGlnuSerTyrrhGlnLysAspGlyAsnGlu 866
      :||| :||| :||| :|||
3917 AGATATTAATCTTGAAGCCAGCAATTTCTGATGAGAAACATCAACAA 3966
      :||| :||| :||| :|||
867 AlaLeuHisGlnLeuIleAspArgMetAsnValLeuGlnLupheThrAl 883
      :||| :||| :||| :|||
3967 AAATTTGCTCATTTTATGATTCTCTTAGTAACCTGGAAGAA... 4008
      :||| :||| :||| :|||
883 AlaLeuMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeu 900
      :||| :||| :||| :|||
4009 .TTGATCTTCTCTACTGCGGATGGAATTTATGAGTGCCCAACTGACAA 4057
      :||| :||| :||| :|||
900 euLysHisLeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsn 916
      :||| :||| :||| :|||
4058 TCCAGCAGCTGCACAGCTTATGTTCTCCGAGTCCCTCTCAATTTTTCAG 4107
      :||| :||| :||| :|||
917 TrpArgLeuThrAspThrGlnIleArgIleLeuGly..... 928
      :||| :||| :||| :|||
4108 ...ACTTGAATGATGACAGCGGTGTGGAATTTGTGAGTAGTGTTCGA 4154
      :||| :||| :||| :|||
929 .....Alap 930
      :||| :||| :||| :|||
4155 GCTTGATGGAAGCCAGTGTATAGCCAAAGCTTTCTGTGCAACATGCTCT 4204
      :||| :||| :||| :|||
930 hePheGlyLysAsnProLeuLysAsnPheGln.....Gln 941
      :||| :||| :||| :|||
4205 ATGTAACATTTGGCCCTTAGAAATTTTCAACCCGCTTCTCTCATTTTCA 4254
      :||| :||| :||| :|||

```

```

942 LeuAsnLeuAlaGlyAsnArgValSerSerAspGlyTrpLeuAlaPhe 958
|||||:|||||:|||||
4255 CATCATACTGTTCTCTTAGTGTCTTCTGTGGAATTTAGGGCATTCG 4304
958 tGlyValPheGluAsnLeuLysGlnLeuValPheAspPheSerThr. 974
:::|||||:|||||:|||||:|||||:|||||
4305 GTCAGATTGGAGTACAAAAGAGTCTCCATTGTGGATATACAAAGCCC 4354
975 .... LysGluPheLeuProAspProAlaLeu 983
:::|||||:|||||
4355 TCAAATCTGCTTCTTGCCACCTGCTGTTTA 4386

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA1996.DAT: AAT30092
seq_documentation_block:
ID AAT30092 standard; cDNA: 5502 BP.
XX
AC AAT30092:
XX
DT 30-DEC-1996 (first entry)
XX
DE Neuronal apoptosis inhibiting protein coding sequence.
XX
KW Neuronal apoptosis inhibiting protein; human; NAIP; chromosome 5q13; YAC;
NM yeast artificial chromosome; spinal muscular atrophy; mammalian cell;
AM autosomal recessive; neurodegenerative disorder; alpha motor neuron; SMA;
KW spinal cord; proximal voluntary muscle; therapy; apoptotic mechanism; ss.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FT CDS 396..4094
FT /tag= a
FT /product= neuronal apoptosis inhibiting protein
XX
XX WO9612016-A1.
XX
XX 25-APR-1996.
XX
XX 17-OCT-1995; 95WO-CA00581.
XX
XX 19-DEC-1994; 94CA-2138425.
XX
XX 18-OCT-1994; 94GB-0021019.
XX
XX (SHKJ) RES DEV CORP JAPAN.
XX
XX (OTOT-) UNIV OTTAWA.
XX
XX Ikeda J, Korneluk RG, Mackenzie AE, Mahadevan MS;
XX McLean M, Roy N;
XX
XX WPI: 1996-222003/22.
XX P-PDB: AAR89217.
XX
XX Neuronal apoptosis inhibitor protein gene - used to develop prods.
XX for use in the diagnosis and therapy of spinal muscular atrophy
XX
XX Claim 4; Page 64-67; 113pp; English.
XX
XX This sequence represents the cDNA sequence for the human neuronal
XX apoptosis inhibitor protein (NAIP). This sequence was found on a region
XX of the human chromosome 5q13. This sequence was isolated from a yeast
XX artificial chromosome (YAC) contig containing the D5S435-D5S112 interval
XX of the chromosome 5q13. Mutations in this gene, are causative of spinal
XX muscular atrophy (SMA) types I, II, and III. SMAs are a group of
XX autosomal recessive, neurodegenerative disorders. SMAs are classified
XX into the three types based upon the age of onset (with type I being the
XX severest form with the earliest age of onset). All three types are
XX characterised by the degeneration of the alpha motor neurons of the
XX spinal cord manifesting as weakness and wasting of the proximal voluntary
XX muscles. The most common mutations of this sequence are thought to be
XX deletions of exons 5 and 6, and reductions in the copy number of the
XX gene. This gene, (and primers and probes based on this gene) can be used

```

```

CC for the diagnosis of SMA, and for directing the formulation of
CC conventional and genetic therapies for SMA. Identification of genes
CC showing homology with the NAIP locus, and proteins that interact with
CC NAIP can be used in the elucidation of apoptotic mechanisms in mammalian
CC cells.
XX
SQ Sequence 5502 BP; 1558 A; 1229 C; 1204 G; 1511 T; 0 other;

alignment_scores:
Quality: 561.50 Length: 937
Ratio: 1.123 Gaps: 33
Percent Similarity: 53.362 Percent Identity: 24.120

alignment_block:
US-09-697-089-2 x AAT30092 ..

Align seg 1/1 to: AAT30092 from: 1 to: 5502

69 PheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAspLeuAs 85
|||||:|||||:|||||:|||||:|||||:
1425 TTCTCCAAAATATGAGATCC...TCTGGGAAGTGAAGTCCAGACCTTCA 1471
85 nGlyGln.....SerLeuPheHisGlnThrSerGluGlyAspL 98
|||||:|||||:|||||:|||||:|||||:
1472 GAGCCGCTGTTGACTTGTGATTTACTGGAACCAACCAAGTGAACCAATC 1521
98 euAspAsp.....
|||||:|||||
1522 TTGAAGATTCAATAGCAGTGTGCTTATAGTCCAGAAATGACACAGAGT 1571
101 .....LeuAlaGlnAspLeuLysAs 107
1572 GAAGCCCACTGTTTCAAGAGGCAAAAGATCTGAATGAGCAGCTGAGAGC 1621
107 PLeuTrpHisThrProSerPheLeuAsnPhe..... 117
|||||:|||||:|||||:|||||:|||||:
1622 AGCTTATACACAGCCGACGTTCCGCCACATGCTTGGTGAATATCTCTT 1671
118 .....TyrProLeuGlyGluAspIleAspIlePhe 128
|||||:|||||:|||||:|||||:|||||:
1672 CCGATCTGGCCAGCAGCAGCAGCTGCTGGCTGTGATCTGTATTT..... 1715
129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnH 145
|||||:|||||:|||||:|||||:|||||:
1716 ..GCTCAAAACACATCAGCAACCTGTG..... 1742
145 sHisHisArgValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeu 162
|||||:|||||:|||||:|||||:|||||:
1743 .....CAAGAACCTCTGCTGCTGCTGCTGAGGCTTTGGCAACTTGA 1782
162 InSerProCysIleIleGluGlyGlnSerGlyLysGlyLysSerThrLeu 178
|||||:|||||:|||||:|||||:|||||:
1783 ACTGTGTCATGTGTGTGGAGGCTGAAGCTGAAGGAAAGCAGCTCTC 1832
179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuTh 195
|||||:|||||:|||||:|||||:|||||:
1833 CTGAAGAAATAGCTTTCTGTGGGCACTGTGATCTGCCCTGTAA 1882
195 LysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
|||||:|||||:|||||:|||||:|||||:
1883 CAGGTTCAGGCTGTTTCTACTCTTCTGCTTGAATTCACCAAGACAGACG 1932
210 lGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226
|||||:|||||:|||||:|||||:|||||:
1933 AGGGCTGCGCAGTATCATCTGTGACCGCTCTGAGAAAGAGAGACT 1982
227 lIeaArgLysGlnThrPheMetAlaMetLeuLysLeuArgGlnArgVa 243
|||||:|||||:|||||:|||||:|||||:
1993 GTTACTGAATGTGCATGAGAACATTTATCCAGCAGTTAAAGATCAAGT 2032
243 lLeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnsnc 258
|||||:|||||:|||||:|||||:|||||:

```

2033 CTTATTCCTTTAGATGACTACAAAGAAATATGTCATCCCTCAA... 2078
 258 yspFrogluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
 2079GTCATAGGAAACGATTCATCAAAAAACCACTATCCCGACCTGC 2123
 275 ValIleValThrThrThrThrGluCysLeuArgHisIleArgGlnPheG1 291
 2124 CTAATGATTCCTGCTCCGACAAACAGGCGACGACATCCGCCATACCT 2173
 291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla1 308
 2174 AGAGACCATTCCTAGAGATCCAGACATTCCTTTTAAATACGTCTGTGA 2223
 308 euIleArgGluValIleLeuLysGluLeuAla....GluGlyLeuLeu 322
 2224 TATTAGCGAAGCTCTTTACATATATGATGCTGCTGCGAAGCTTATG 2273
 323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProLe 339
 2274 GTTATCTTTGAAAGAACCAAGTTTCAGAGATACAGAAATCCTCT 2323
 339 upIleValIleThrCysAlaIleGlnMetGlyGluSerGluPheHis 356
 2324 CTTTGTGCGCGCATCTGTGCTCATTTGGTTCACTATCCTTTGACCCAT 2373
 356 eRhIsthrGlnThrThrLeuPheHisThrPheTyrAspLeuIleGln 372
 2374 CCTTGATGATGTCGCTGTTTCAAGCTATATGAAACGCTTCTCTTA 2423
 373 LysAsnLysHisLysHisLysGlyValAlaAlaSerAspHeIleArg... 388
 2424 AGGAACAAA.....GCGACAGCTGAATTCCAAGC 2455
 389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405
 2456 AACTGTCTCTCTCTGTCGTGACGTGCGCTTGAAGGCTTTTTCATGTT 2505
 405 yspPheAspPheGluLeuGlnAspVal....SerSerValAsnGluAsp 419
 2506 GCTTTGATTAATGATGATGATCTCCAGACAGCGGTGATGAGAT 2555
 420 ValLeuLeuThrThrGlyLeuLeuGlyLysTyrThrAlaGlnArgPheLys 436
 2556 GAAGATCTAACCATGCTGTCGATGAGCAAAATTTACACCCAGAACACTAG 2605
 436 sProlYstYrLysPhePheHisLysSerPheGlnIuTyrThrAlaGly 453
 2606 ACCATCTACCGGTTTTTAAGTCTGCTCCCAAGAAATTTCTCGGGGA 2655
 453 rgrArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
 2656 TGAGGCTGATTTGAACAATCACTACCTACCATGATGACTGTAACGCG 2705
 470 GlyAsnGlyThrLeuGlnLysMetValSerIleSerAspIleThrSerth 486
 2706 GGACTGATATTTGAACAATCACTACCTACCATGATGACTGTAACGCG 2755
 486 rTySerSerLeuLeuArgTyrThrCysGlySerSerValGluIleAlaThra 503
 2756 CTACAAACAATTTTGAACATATGTC....TCCAGCTCCCTTCAACAA 2799
 503 rglAla.....ValMetLysHisLeuAlaAlaValTyrGlnHisGly 516
 2800 AACGACAGGCCCAAAATGTGTCATATTGCTCATTTAGTGAGTAAC... 2846
 517 CysLeuLeuGlyLeuSerIleAlaIleLysArgProLeuThrPargGlnGluSe 533
 2847AAAGAGTC 2854
 533 rLeuGlnSerValLysAsn..... 539
 2855 ATTGGAGAAATATATCTGAAATAATGATGACTTAAGCACCAGCAGAAA 2904

540 ..ThrThrGluGlnGluIleLeuLysAlaIle.....AsnIle 551
 2905 TTTCACGACGATGACATTTACTTAGGGGATTTGTGCAAAATTTGTCCAA 2954
 552 AsnSerPheValGluCysGlyIleHisLeu..... 561
 2955 GCTTACTTTCAATGATTTGAGAACATTTACTGTGTCCTCGAAGAAC 3004
 562 ...TyrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheGlu 577
 3005 TGCTTATCAAGCACACT...GTTGCTGGCTGTTCATTTGTTTTC 3051
 577 lAspPheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsp 593
 3052 AATTCCTTCAAGGAGAACACTGACTTTGGGTGCGCTTAACCTTA...CAG 3098
 594 TyrLeuPheAsp..... 597
 3099 TACTTTTGCACCAACCCAGAAAGCTGTCTCATTTGTGAGGACATCCACTT 3148
 598PhePheGluHisLeuP 603
 3149 CTCATATCGAGAAATATAGACATCACCCAGACACATTTTCACTTCTGG 3198
 603 roAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
 3199 AAACATGTTTTGACAAATACAGAGTGCACACTATATGATCAGACTATGCT 3248
 618 GlyGly.....AlaMetAlaSerThrGlu..... 625
 3249 TCTGCTTTGAACTTATGATGATGATGGAGCGAAATTTAGCTGAAGAAA 3298
 626LysAlaAlaGluAspT 631
 3299 GGATATCTTAAGAGCTATATGATATGACAGCGAGCGCATTCACAGACC 3348
 631 hrgIlyIleHisMetGluGluAlaProGluThrTyr...IleProSer 646
 3349 TTAGTACTGCTATTTGAAACTTCTCCAAAGCAGTACAGATATCCC... 3395
 647 ArgAlaValSerLeuPhePheAsnThrLysGlnGluPheArgThrLeuG1 663
 3396TGTCTAGA 3403
 663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrL 680
 3404 AGTCGATGTGATGATATTTGATGTTGAGGCCAGAGATATGCTGACATTC 3453
 680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
 3454 TTAATGACAGTTTCTCAGCTTCACAGCGCATCGAATCCATTTAAACCA 3503
 697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
 3504 AGCAGAGGCTTTATAGAAAGCATCCGCCAGCTTGTAGCTGTCTAGGCG 3553
 713 ..AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAspG 729
 3554 CTCTGTACCAAGTCCTCATTAAGCAAGTTGGAAGTACAGCGACCGCAAC 3603
 729 lArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
 3604 AGGAACGCTTTCACCCCTGCTCCCTGGAATCTGTAAGCTCAGGG 3653
 745AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
 3654 ACAATCCAGTCCACAGACCAAACTTCTTCTTAAT....CTGATTAAGTT 3697
 759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet.. 775
 3698 CTTGTGCTGTAAGAAACACTGTCTGTGATCTGAGAGGCAATATTAATGTTT 3747


```

775 ..... 775
3748 TTTCAGTCAATTCCTGAGAAATTTCCAACTTCACCATATGGAATTA 3797
776 ..... 787
3798 TTGATCAAAATTTTCAGTATGATTCCTCCAACTGATGTCAGTTT 3847
787 uLysAsnLeuLysLysMetCysLeuPheHisLeuThrHisLeu..... 801
3848 GCCAAATTTTATTTCTCTGAAGATTTAAATCTTGAGGCCACCAATTC 3897
802 .....SerAspIleGlyGluGlyMetAspTyrIleValLysSerLeuSer 816
3898 CTGATGAGGAACATCAGAAAAATTTGCTACATTTTAGTTTCTCTAGT 3947
817 SerGluProCysAspLeuGluGluIleGlnLeuValSerCysLeuSe 833
3948 .....AACCTGGAGAAATGATCTCTCTACTGGGATGGAAT 3985
833 rAlaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysL 850
3986 TTATCAGATGGCCAAACTGATCATCCAGCATGTCAGCAGCTTCATTTGTC 4035
850 euserIleLeu 853
4036 TCCGAGTCTC 4046

```

seq_name: /SIDS2/gcdata/geneseq/geneseq/NA197.DAT:AA71263

seq_documentation_block:

ID AA71263 standard; cDNA; 5502 BP.

AC AA71263:

DT 06-OCT-1997 (first entry)

DE Neuronal apoptosis inhibitor protein (NAIP) cDNA (no exon 14a).

KW Neuronal apoptosis inhibitor protein; NAIP; diagnosis;

KW gene therapy; cancer; AIDS; amyotrophic lateral sclerosis;

KW spinal muscular atrophy; ss.

OS Homo sapiens.

PN W09726331-A2.

PD 24-JUL-1997.

PF 17-JAN-1997; 97WO-IB00142.

PR 19-JAN-1996; 96GB-0001108.

XX (UYOT-) UNIV OTTAWA.

XX Korneiluk RG, Mackenzie AE, Robertson G, Roy N, Tamai K;

XX WPI; 1997-38535/35.

XX New neuronal inhibitor of apoptosis - useful for diagnosing and

XX treating, e.g. cancer, AIDS or amyotrophic lateral sclerosis

XX Example 4; Fig 5A-L; 102pp; English.

CC A new neuronal apoptosis inhibitor protein (NAIP) cDNA clone

CC (AA71263) was isolated from a human foetal spinal cord cDNA library

CC by probing with the genomic insert in cosmid 25086, containing a

CC CAPT locus. An additional coding sequence (AA71264), including a

CC exon 14a (see also AA71265 and AA71266), was subsequently obd.

CC The NAIP DNA sequence including exon 14a appears to be a

CC predominant gene isoform which is not deleted or mutated in spinal

CC muscular atrophy (SMA) patients. The NAIP gene was mapped to

CC 5q13.1. NAIP (see also AAW20032 and AAW20033) is a negative regulator

CC of apoptosis, partic. neuronal apoptosis and, when deficient or
 CC absent, contributes to neurodegenerative phenotypes such as SMA and
 CC amyotrophic lateral sclerosis.

XX SQ Sequence 5502 BP; 1560 A; 1226 C; 1206 G; 1510 T; 0 other;

alignment_scores:

Quality:	546.50	Length:	950
Ratio:	1.078	Gaps:	38
Percent Similarity:	53.368	Percent Identity:	24.000

alignment_block:

US-09-697-089-2 x AA71263

Align seg 1/1 to: AA71263 from: 1 to: 5502

69 PheLeuLysSerLeuLysGluTrpAsnTyrProLeuPheGlnAspLeuAs 85

1425 TTCTCTCAAAATATGAAATGTC...TCTCGGAGTACATCCAGACCTTCA 1471

85 nGlyGln.....SerLeuPheHisGlnThrSerGluGlyAspL 98

1472 GAGCCCTGGTGAACCTTGTGAATTAATGGAACCAAGCAAGCAATC 1521

98 euAspAsp..... 100

1522 TTGAAGATTCATATAGCAGTTGCTATATAGTCCAGAAATGCGACAGGT 1571

101LeuAlaGlnAspLeuLysAs 107

1572 GAAGCCAGTGGTTTCAAGGCGCAAAATCTGAATAGACAGCTGAGAC 1621

107 pleuTYrHisThrProSerPheLeuAsnPhe..... 117

1622 AGCTTATACACAGCGCATGTTCCGCCACATGCTTTCCTGATATCTCT 1671

118TyrProLeuGlyGlnAspIleAspIlePhe 128

1672 CCGATTCGGCCAGCGACCACTTGGCTGGCTGATGATCTATAT..... 1715

129 AsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnH 145

1716 ...GCTTCAAAACATCAGCAAACTGTG..... 1742

145 SHSHSHArGValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuG 162

1743CAAGAACCTCTGCTGCTGCTGAGGCTTTGGCAACTTGA 1782

162 InSerProCysIleIleGluGlyGluSerGlyLysGlyLysSerThrLeu 178

1783 ACTCTGTCAATGTGTGTGAAGGTGAAGCTGGAAGTGAAGAGCGGTCTC 1832

179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuTh 195

1833 CTGAAGAAATAGCTTTTCTGTGGCATCTGATGTCGCCCTGTATAA 1882

195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGlnG 210

1883 CAGGTTCCAGCTGGTTTCTACTCTCCCTTGTAGTTCACACAGCAGAGC 1932

210 LysGlyLeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThr 226

1933 AGGGCTGGCCAGTATCATCTGTGACCACTCTTAAAGAAAGAGATCT 1982

227 IleArgLysGlnThrPheMetAlaMetLeuLysLeuArgGlnArgVa 243

1983 GTTACTGAATGATGATGGAAGCAATTAATTCACAGCTTAAGCAATCAGCT 2032

243 IleuPheLeuAspGlyTyrAsnGluPhe.....LysProGlnAsn 258

2033 CTATATCTTTAGATGAGTACAAAGAAATATGTTCAATCTCTCAA.... 2078


```
,
3748 TTTCA.....GTCAATTCCTGAAGAAATTTCCAAC..... 3776
793 MetCysLeuPheHisLeuThrHisLeuSerAspIleGlyGluGlyMetAs 809
3777 .....TTCCAC.....CATATGGA 3790
809 pTyrlIleValIysSerLeuSerSerGlu.....ProCysAspLeuGluG 824
3791 GAAATTATTGATCCAAATTTTCAGCTGAGATGATGATCCTTCCAACACTAGT.. 3838
824 IuIleGlnLeuValSerCysCysLeuSerAlaAsnAlaValIysIleLeu 840
3839 .....AATGCC..... 3844
841 AlaGlnAsnLeuHisAsnLeuValIysLeuSerIleLeuAspLeu...Se 856
3845 .....AGTTGCCAAATTTTATTTCTCTGAGAGATATTAAATCTTGAAAG 3888
856 rGluAsnTyrlLeuGluIulysAspGlyAsnGluAlaLeuHisGluLeuIleA 873
3889 CCAGCAATTTCTCTGATGAGAGAAACATCAGAAAAAATTTGCTACATTTTAG 3938
873 sPArgMetAsnValLeuGluGlnLeuThrAlaLeuMetLeuProTPrGly 889
3939 GTTCTCTAGTAACCTGGAAGAA.....TTGATCCTCTCTACTGSG 3979
```

THIS PAGE BLANK (USPTO)

OM of: US-09-697-089-2 to: EST:* out_format: pfs
Date: Mar 26, 2002 2:05 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=firmest p2n.model -DEV=xih
-O/cgnt2_1/USPTO.spool/US09697089/runatc_25032002_105137_11138/app_query.fasta_1.1097
-DB=EST -QMT=fastap -SUFFIX=est -GAPOP=12.000 -GAEXT=4.500
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsun62 -TRANS=human40.cit
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09697089_ECGNT_1_4971
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLRPY -WAIT -THREADS=1

Search information block:

Query: US-09-697-089-2
Query length: 1024
Database: EST*
Database sequences: 11351937
Database length: 107921985
Search time (sec): 1177.210000

score list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
gb_gss:AQ0309404	-	952.00	1809.29	1.3e-91	552	AQ0309404 CITBI-E1-2528J13.TF CT
gb_gss:A2360053	-	811.00	1537.20	1.9e-76	602	A2360053 IM0103H1F Mouse 10kb
gb_est1:AV1919179	-	801.00	1520.37	1.6e-75	480	AV1919179 AV719179 GLC Homo sapi
gb_est1:AV656315	-	593.00	1127.07	2.2e-53	371	AV656315 AV656315 GLC Homo sapi
gb_est1:AI263324	-	585.00	1107.88	1.5e-52	364	AI263324 RST529913 X1 NCI CGAP-P
gb_est2:BG210375	-	475.50	894.46	1.2e-40	476	BG210375 RST29913 Altherys RAGE
gb_gss:AQ624020	-	445.50	836.90	1.9e-37	470	AQ624020 HS_5378_B2_C12_Sp6E RH
gb_est1:AW337918	-	377.00	711.38	1.8e-30	261	AW337918 he12h11.X1 NCI CGAP-CM
gb_gss:AQ088169	-	323.00	602.91	2.0e-24	404	AQ088169 HS_2161_B1_A01_T7C CIT
gb_gss:AQ030928	-	298.00	551.50	1.5e-21	553	AQ030928 RST529913 X1 NCI CGAP-P
gb_gss:AQ088169	-	260.00	488.22	5.0e-18	219	AQ088169 RST529913 X1 NCI CGAP-P
gb_gss:AQ030928	-	255.00	467.43	7.1e-17	630	AQ030928 RST529913 X1 NCI CGAP-P
gb_est2:BF903662	-	248.50	466.66	1.0e-16	251	BF903662 CIT-HSP-2372C1.TF CIT
gb_gss:A2614134	-	244.00	446.36	1.1e-15	625	A2614134 IM0442M1R Mouse 10kb
gb_est1:AW118826	-	222.00	406.22	1.8e-13	509	AW118826 ha21e11.X1 NCI CGAP-K1
gb_est2:BF207840	-	205.50	369.19	2.1e-11	840	BF207840 601862546F1 NIH_MGC_53
gb_gss:A2720059	-	189.50	339.93	9.0e-10	728	A2720059 RST529913 X1 NCI CGAP-P
gb_gss:A2360053	-	187.00	338.17	1.1e-09	546	A2360053 IM0107N03R Mouse 10kb
gb_est2:H25984	-	174.00	316.14	1.9e-08	412	H25984 Y15607.X1 Soares breast
gb_est1:AI023795	-	174.00	314.12	2.5e-08	499	AI023795 ox08D03.X1 Soares_feta
gb_gss:A2805292	-	171.50	313.30	2.7e-08	342	A2805292 2M0066B12R Mouse 10kb
gb_hlc:AK017656	-	163.00	280.07	2.0e-06	1688	AK017656 Mus musculus 8 days e
gb_hlc:AK017780	-	163.00	279.25	2.2e-06	1824	AK017780 Mus musculus 8 days e
gb_gss:AK019545	-	156.00	258.16	3.2e-05	3750	AK019545 Mus musculus adult ma
gb_gss:A2762115	-	143.50	252.37	6.8e-05	668	A2762115 IM0556N1R Mouse 10kb
gb_hlc:AK015024	-	143.00	239.59	0.0004	2043	AK015024 Mus musculus adult ma
gb_hlc:AK009342	-	142.50	243.21	0.0002	1324	AK009342 Mus musculus adult ma
gb_hlc:AK009342	-	141.50	237.40	0.0005	1913	AK009342 Mus musculus adult ma
gb_est2:BG448903	-	141.00	248.25	0.0001	626	BG448903 NF00504E1N1043 Insec
gb_est1:AU132668	-	139.50	247.00	0.0003	861	AU132668 NF00504E1N1043 Insec
gb_hlc:AK019775	-	139.00	230.53	0.0011	2324	AK019775 Mus musculus adult ma
gb_hlc:AK017814	-	138.00	230.06	0.0012	2026	AK017814 Mus musculus 8 days e
gb_hlc:AK011726	-	138.00	229.70	0.0012	2085	AK011726 Mus musculus 10 days e
gb_gss:AQ0967565	-	136.50	237.87	0.0004	737	AQ0967565 LERIS91TF LERG Arabid
gb_est1:AL138332	-	136.00	237.59	0.0005	691	AL138332 DKEF762N043.X1 762 (S
gb_hlc:AK017842	-	135.00	226.17	0.0020	1696	AK017842 Mus musculus 8 days e
gb_gss:A2547972	-	134.00	231.54	0.0010	851	A2547972 EMTD777TF Entamoeba h
gb_hlc:AF054176	-	133.00	219.48	0.0046	2218	AF054176 Homo sapiens angioten
gb_est2:BE83335	-	132.00	236.11	0.0005	384	BE83335 OV3-OT0055-280600-250-
gb_hlc:AK014400	-	132.00	211.29	0.0132	4013	AK014400 Mus musculus adult ma

gb_est2:BF499764 + 131.00 228.40 0.0015 664 BF499764 AT14482.5p1rme AT D
gb_est2:BG684008 + 130.50 225.40 0.0022 805 BG684008 602635613F1 NIH_MGC
gb_est2:BI409867 + 130.00 222.34 0.0032 981 BI409867 602962204F1 NCI_CGA
gb_gss:CNS02NVI - 129.50 220.78 0.0039 1039 AL205894 Tetraodon nigrovir
gb_est2:BG994334 - 129.00 225.87 0.0020 586 BG994334 PM0-HT1166-130201-0

seq_name: gb_gss:AQ0309404

seq_documentation_block:

LOCUS AQ0309404 552 bp DNA 22-DEC-1998
DEFINITION CITBI-E1-2528J13.TF CITBI-E1 Homo sapiens genomic clone 2528J13,
DNA sequence.

ACCESSION AQ0309404
VERSION AQ0309404.1 GI:4041438

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 552)
Berry, K., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Ventner, J.C., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.

AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.

TITLE Map Building
Use of a random human BAC End Sequence Database for Sequence-Ready
Mapping

Other-GSS: CITBI-E1-2528J13.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208

EMAIL: hncetlgr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.figr.org/tbdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
Location/Qualifiers
1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2528J13"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

BASE COUNT 144 a 121 c 130 g 157 t
ORIGIN

alignment_scores:
Quality: 952.00 Length: 183
Ratio: 5.202 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-697-089-2 x AQ0309404/rev ..

Align seg 1/1 to reverse of: AQ0309404 from: 1 to: 552

477 MetValSerIleSerAspIleThrSerThrTySerSerLeuLeuArgTty 493
|||||
550 ATGGTTTCATTCGACATTCATTCATTCATTCATTCATTCATTCATTCGGA 501

493 rthrcysgyserserValGluAlaThrArgAlaValMetysHisLeuA 510
|||||
500 CACCTGTGGTCATCTGTGGAAGCACAGGCTGTATGAAGCACCTCG 451

510 lAlaValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAlaTysArg 526
|||||

```

450 CAGCAGTGTATCAACACGCGCTCTCGACTTTCATCGCCACAGAG 401
527 ProleuTripargIngluSerleuGlnSerValIysasnThrInglu 543
|||||
400 CGCTCTGGAGACAGCAATCTTTGCAAAAGTGTAACCAACACTGAGCA 351
543 ngIuIleuIysAlaIleasnIleasnSerPheValGluCysGlyIleH 560
|||||
350 AGAAATCTCAAAAGCCATAACATCAATTCCTTTGTAGAGTGCGCATCC 301
560 IIsuIyrIngluIserThrSerIysSerIaleuSerGlnIuPheGlu 576
|||||
300 ATTATATTCAGAGAGTACATCCAAATCAGCCCTGAGCCCAAGATTGAA 251
577 AlaPhePheGlnGlySerSerleuTyrlIeasnSerGlyasnIleProas 593
|||||
250 GCATTCTTTCAGAGTAAAGCTTATATATCACTACAGGAAACATCCCGA 201
593 pTyrlLeuPheAspPhePheGluHisIleuProasnCysAlaSerAlaIeuA 610
|||||
200 TTACTTATTGACTCTTTGACATTTGCCCAATTTGCAAGTCTCTCG 151
610 sPheIleIysIleuAspPheTyrlGlyIleAlaMetAlaIserTrpGluIys 626
|||||
150 ACTTCATTTAACTGAGCTTTATGAGGAGCTATGGCTTATGGAAGAAG 101
627 AlaAlaGluAspThrGlyGlyIleHisMetGluGluAlaProGluThrTy 643
|||||
100 GCTGAGAGACACACAGGTGGAATCCACATGGAAGAGCCCAAGAACTTA 51
643 rIleProSerArgAlaValSerleuPhePheasnTrpIysGlnIuPhe 659
|||||
50 CATTCACGACGAGGCTGTATCTTTGTTCTTCACACTGGAACAGCAATTC 2
seq_name: gb_gss:AZ360053

```

```

seq_documentation_block:
LOCUS      AZ360053      602 bp      DNA      GSS      02-OCT-2000
DEFINITION 1M0103H11F Mouse 10kb plasmid UGCC1M library Mus musculus genomic
clone UGCC1M0103H11 F, DNA sequence.
ACCESSION  AZ360053
VERSION     AZ360053.1 GI:10473753
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 602)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
            and Wright,D.,Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
TITLE      unpublished (2000)
JOURNAL    Contact: Robert B. Weiss
COMMENT    University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert length: 10000 Std Error: 0.00
            Plate: 0103 row: H column: 11
            Seq primer: GGTGTAAACGACGCGCCACT
            Class: plasmid ends
            High quality sequence stop: 602.
            Location/Qualifiers
                1..602
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"

```

```

/clone="UGCC1M0103H11"
/clone_11b="Mouse 10kb plasmid UGCC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pMD42ny. Purified genomic DNA from M.
musculus C57BL/6J (male). Was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using prepared agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      139 a      153 c      139 g      171 t
ORIGIN

```

```

alignment_scores:
    Quality:      811.00      Length:      195
    Ratio:        4.481      Gaps:        0
    Percent Similarity: 92.821      Percent Identity: 79.487

```

alignment_block:

us-09-697-089-2 x AZ360053/rev ..

Align seg 1/1 to reverse of: AZ360053 from: 1 to: 602

```

382 ALAAlaSerAspPheIleArgSerleuAspHisCysGlyAspIleAla 398
|||||
587 GCTTCAGGTATTTGGCAGAGAGCGCTGACTGATGGAACGCGGCGCT 538
398 uGluGlyValPheSerHisIysPheAspPheGluIleGlnAspValSers 415
|||||
537 AGAAGGTGTGTCGCCCAAAATTTGATTTGAACCCGACGATGGGTCA 488
415 eValAsnGluAspValIleuThrThrGlyLeuIleuCysIysTyThr 431
|||||
487 GCATGAACGAGAGAGCTCTGTCGACATAGGCGCTCTGTAACTACACA 438
432 AlaGlnArgPheIysProIysTyrlIysPhePheHisIysSerPheGln 448
|||||
437 GCTCAGAGCGCTGAAGCCACGATTAATTTCTTCATTAATCATTTACGA 388
448 uTyThrAlaGlyArgArgIleuSerSerleuIleuThrSerHisGluProG 465
|||||
387 GTACACGCGAGGTGCGAGACTCAGACGTTTGTGACCTCCAAAGAGAG 338
387 IuGluValThrIysGlyAsnGlyTyrlIeGlnIuIysMetValSerIleSer 481
|||||
337 AGGAGGTGACCAAGGAGACACTTAACAAATGGTTTCATCTCT 288
482 AspIleThrSerThrTyrlSerSerleuIleuArgTyrlThrCysGlySer 498
|||||
287 GACATCAATCCCTATATGCAATCTGCTCTACACGATGGGTGCTC 238
498 rValGluAlaThrArgAlaValMetIysHisIleuAlaIleValTyrlInH 515
|||||
237 CACGAGCAACCAACGAGGCGGTCAAGAGCACCTTGCAATGTTATAGC 188
515 IsgIyCysLeuIleuGlyIleuSerIleAlaIysArgProIeTriParGln 531
|||||
187 ACGGACCTCAACAGAGACTTTCAGTCAACCAAGAGCGCTCTCTGAGGAG 138

```

seq_name	gb_estl	AV656315
seq_documentation_block:		
LOCUS	AV656315	371 bp mRNA
DEFINITION	AV656315	GLC Homo sapiens cDNA clone GJCBA10 3', mRNA sequence.
ACCESSION	AV656315	
VERSION	AV656315.1	GI:9877329
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 371) Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.	
TITLE	Homo sapiens cDNA clone	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhongjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel.: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.	
FEATURES	Location/Qualifiers	
SOURCE	1..371 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="GJCBA10" /clone_id="GJC" /tissue_type="corresponding non cancerous liver tissue" /dev_stage="Adult" /lab_host="SOLR" /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	112 a 85 c 91 g 82 t 1 others	
ORIGIN		

alignment_scores: Length: 141
 Quality: 593.00 Gaps: 1
 Ratio: 4.861
 Percent Similarity: 86.525 Percent Identity: 86.525

alignment_block:
 US-09-697-089-2 x AV656315 ..

Align seq 1/1 to: AV656315 from: 1 to: 371

```

625 GLUYSALALAGLUASPTGRLGTYLLEHISMETGUGLUALPROGL 641
      |||||||
1  GAAAGGCTGCAGAGACACAGGTGGAATCCACATGGAGAGGCCCGCA 50
      |||||||
641 UTHYTLLEPROSERARGALALASERLEUPHEASNTIRPLYSGLNG 658
      |||||||
51 AACCTACATTCGCCAGAGGCGCTGATCTTTCTTCAACCGAAGCAGG 100
      |||||||
658 LUPHEATGTHLEUGLUVALTHLEUARGLSPHEASERLYSLEUASNL 674
      |||||||
101 AATTCAGGACCTGAGGTGACACTCCGGGATTTCCAGCAAGTTGAAT 150
      |||||||
675 GLNASPLIETHYTLLEUGLYLLEPHESESERIALTHRSERLEUAR 691
      |||||||
151 CAGATATTCAGATATCTGGGGAAATATTCAGCTCTGCCACAGCCTCAG 200
      |||||||
691 GLEUGLILELYSARGCYSLAGLYVALAGLYSERLEUSERLEUVAL 708
      |||||||
201 GCTCCAAATGAAGAGATGTGCTGGTGGCTGGAGCCTCAGTTGGGCC 250
      |||||||
708 EUSERTHCYSLYSASNILETYRSERLEUETVALGLUALASERPROLE 724
      |||||||
251 TCACACCTGTAAGACATTTATCTCTCATGCTGGAGCCAGTCCCTC 300
      |||||||
725 THRILEGLUASPRGLUARGHISILETHRSERVALTHRSASNLEUL 741
      |||||||
301 ACCCTAAGAGTGAGAGCAGCATCATCTGTAAACAAACCTG..... 342
      |||||||
741 USERILEHISASPLEUGLNASNGLARGLEUPROGLYLEUTHRASPS 758
      |||||||
343 .....GACA 346
      |||||||
758 ERLLEUGLYASNLEULYSASNLEU 765
      |||||||
347 GCTTGGGTAACTTGAAGAACCTT 369
      |||||||

```

seq_name: gb_est1:AI263294

documentation_block:

US AI263294 364 bp mRNA EST 03-FEB-1999
 DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
 mRNA sequence.
 ACCESSION AI263294
 VERSION AI263294.1 GI:3871497
 KEYWORDS EST.

SOURCE
 ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 364)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Life Technologies catalog #: 11548-013

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 2146
 Std Error: 0.00
 Seq primer: -40UP from Gibco

FEATURES High quality sequence stop: 364.
 Location/Qualifiers
 source
 1..364

/organism="Homo sapiens"
 /gb_xref="taxon:9606"
 /clone="IMAGE:2005417"
 /clone_1ib="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: pCMV-Sport6; Site:1; Salt:
 Site:2; Note: Cloned unidirectionally. Primer: Oligo df.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"

BASE COUNT 117 a 84 c 55 g 108 t
 ORIGIN

alignment_scores: Length: 114
 Quality: 585.00 Gaps: 0
 Ratio: 5.132
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AI263294/rev ..

Align seq 1/1 to reverse of: AI263294 from: 1 to: 364

```

911 LYSLEUGLYLEULYSASNTRPARGLEUTHRASPTGRLGUILLEARGILELE 927
      |||||||
364 AACCTGGGTTGAAAACCTGGAGACTCAGCATACAGAGATTACAAATTTT 315
      |||||||
927 UGLYALAPHEPHEGLYLSASNPROLEULYSASNPHEDGLNLEUASNL 944
      |||||||
314 AGTGCAATTTTGGAAAACCTCGAAAACCTCGACAGAGTTGAATT 265
      |||||||
944 EUALAGLYASNARGVALSERSERASPGLYTRPLEUALAPHEMEGLYVAL 960
      |||||||
264 TGGGGGAAATCGTGTAGACGATGATGATGCTTCATGATGGTGTA 215
      |||||||
961 PHEGLUASNLEULYSGLNLEUVALPHEPHEASPTHRILYSGLIUPH 977
      |||||||
214 TTTCAGATCTTAAAGCAATTGATGTTTTCACCTTTAGTACTAAAGATT 165
      |||||||
977 ELEUPROASPPROALALEUVALARGLYSLEUSERGLNVALLEUSERYSL 994
      |||||||
164 TCTACCTGATCCAGCATTAAGTACAGAAAACCTTACCAAGTCTTATCCAA 115
      |||||||
994 EUTRPHLEUGLNGLUALARGLEUVALGLYTRPGLNPHASPPASPASP 1010
      |||||||
114 TTAACCTTTCTGCAAGAGCTAGGCTGTGTGGTGCAATTTGATGATGAT 65
      |||||||
1011 ASPLSESERVALILETHNGLYALAPHELYSLEUVALTHRALA 1024
      |||||||
64 GATCTCAGCTGTTATTACAGTGCCTTTTAACTAGTAACTGCT 23
      |||||||

```

seq_name: gb_est2:BG210375

seq_documentation_block:

LOCUS BG210375 476 bp mRNA EST 21-APR-2001
 DEFINITION RST29913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG210375
 VERSION BG210375.1 GI:13732062
 KEYWORDS EST.

SOURCE
 ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 476)
 Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
 Cahn,S., Dahl,T., Thornton,M., Ramchandran,R., Whitlington,J.,
 Lerner,L., Krishnoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
 Veloso,N., Hess,J., Cotthren,K., Lo,K., Offenbacher,J., Danzig,J.
 and Ducar,M.

REFERENCE
 AUTHORS

55 AGATATCATGATATCTGGGGAAAAATATTCACGCTCTGCCACAGCCTCAGGC 104
 692 eucInilelySarGysAlaGlyValAlaGlySerLeuSerLeuValleu 708
 105 TGCAGATTAAGAGATGCTGCTGTGCTGCGAAGCCCTCATTTGCTCTC 154
 709 SerTCysLysAsnIleTySerLeuMetValGlnAlaSerProLeuTh 725
 155 AGCACCCTGAAGAACCATATCTCTCATGTCGTAAGCCAGTCCTCCTCAC 204
 725 rllleGluAspGluArgHisIleThrSerValThrAsnLeuLysThr 742
 205 CATAGAGAGATGAGAGCAGCATCATCTGTACACNACCTGAAACCTTGA 254
 742 eilLeHisAspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSer 758
 255 GTATTTCATGACCTACAGATCAACGCGCTGCCG 286
 759 LeuGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLys 775
 286 286
 775 tAsnGluGlnAspAlaIleLeuLysLeuAlaGlyLeuLysAsnLeuLysL 792
 287GGTATTGTAATAATATCAGTG 305
 792 ysmetCysLeuPheHisLeuThr 799
 306 GTTGTGCTTTGTTCACTTACA 328

seq_name: gb_est1:AW337918

seq_documentation_block:
 LOCUS AW337918 261 bp mRNA EST 31-JAN-2000
 DEFINITION hel2h11.x1 NCI_CGAP_CML Homo sapiens cDNA clone IMAGE:2918853 3',
 mRNA sequence.

ACCESSION AW337918
 VERSION AW337918.1 GI:6834544
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 261)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
 Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 201.

FEATURES

source
 1..261
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2918853"
 /clone_lib="NCI_CGAP_CML1"
 /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
 rearrangement positive, includes both chronic phase and
 myeloid blast crisis"
 /lab_host="DH10B"
 /note="Organ: whole blood; Vector: pCMV-SPORT6; Site:1;
 SalI; Site:2; NotI; Cloned unidirectionally. Primer:"
 oligo dt. Library constructed by Life Technologies."

BASE COUNT 86 a 54 c 38 g 83 t
 ORIGIN
 alignment_scores:
 Quality: 377.00 Length: 74
 Ratio: 5.095 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-697-089-2 x AW337918/rev ..

Align seg 1/1 to reverse of: AW337918 from: 1 to: 261

951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGln 967
 260 AGTATGATGATGCGCTGCGCTTCATGAGGTATTTGAGAGATTTAAGCAT 211
 967 uValPheAspPheSerThrLysGluPheLeuProAspProAlaLeu 984
 210 AGTGTTTTGGACTTGTACTAAAGAAATTTCTACCTGATCCAGCATTAG 161
 984 aAlaGlyLeuSerGlnValLeuSerLysLeuThrPheLeuGlnAla 1000
 160 TCAGAAACTTACGCAAGTGTATCCAAATTTCTTCAGAGAGCT 111
 1001 ArgLeuValGlyTrpGlnPheAspAspAspLeuSerValIleThrG 1017
 110 AGGCTTGTGTTGGTGCGCAATTTGATGATGATGATCTGATTTATACAG 61
 1017 yAlaPheLysLeuValThrAla 1024
 60 TCGTTTAAACTGATGACTGCT 39

seq_name: gb_gss:AQ889169

seq_documentation_block:
 LOCUS AQ889169 404 bp DNA GSS 10-NOV-1999
 DEFINITION HS_2161_B1.A01.TTC CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.

ACCESSION AQ889169
 VERSION AQ889169.1 GI:6345359
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 404)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

JOURNAL
 COMMENT Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
 Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
 Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
 Hood L.
 Sequencing-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

TITLE
 JOURNAL
 MEDLINE
 COMMENT Contact: Mahairas G.G., Wallace J.C., Hood L.
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2161 Row: B column: 1
 Seq primer: 77
 Class: BAC ends
 High quality sequence stop: 404.

FEATURES

source
 1..404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2161 Col=1 Row=B"

/clone.lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 119 a 93 c 70 g 122 t
 ORIGIN

alignment_scores:

Quality: 323.00 Length: 90
 Ratio: 3.892 Gaps: 0
 Percent Similarity: 92.222 Percent Identity: 77.778

alignment_block:

US-09-697-089-2 x AQ889169/rev ..

Align seg 1/1 to reverse of: AQ889169 from: 1 to: 404

```

1 MetasnpheillelysaspasSerArgAlaLeuileGlnArgMetGlyme 17
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
351 GTCAATTTCATAGACGACGATACCCGACG.CTTATTCAGAAATGGCAAT 303
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
17 tthrvalillelysglnlethrapsasphevaltrpasnvalleua 34
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
302 GCCTGTATAGACTACTACAGATGACCTATTGTATGAAATGTATATGA 253
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
34 snarglucguvalasnillelecygsluylsvalgluglnaspala 50
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
252 ATCGAGAGAGAGTAACATCATTTCTGGAGAGGTGGAGCGAGATGCT 203
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
51 AlaargglyllellehismetilleulyslysglysergluSerCys 67
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
202 GTTAGAGGATCATTCACATGATTTGMAAGAGGTTCCAGAGTCCTGTA 153
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
67 nleupheleulysSerleulysglutrpasnryrProleupheglnaspl 84
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
152 CCCCTATGTTAAATCCCTTAAGGAGGTGAACATCTCTATTTTCAGGACT 103
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
84 euaanglyginserleuphe 90
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
102 TGAAGGACAAAGTAAGTAT 83

```

seq_name: gb_gss:AQ320928

seq_documentation_block:

LOCUS AQ320928 553 bp DNA GSS 06-MAY-1999
 DEFINITION RPC111-93C9.TV RPC1-11 Homo sapiens genomic clone RPC1-11-93C9, DNA
 sequence.

ACCESSION AQ320928
 VERSION AQ320928.1 GI:4053662
 KEYWORDS GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 553)
 Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
 Berry K., Granger D., Suh E., Wible C., de Jong P. and Venter J.C.
 Use of human BAC End Sequences for Sequence-Ready Map Building

TITLE

Unpublished (1998)

JOURNAL

Other GSSs: RPC111-93C9.TV

COMMENT

Contact: Shaying Zhao, William Niernan, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetlgr.org

Clones are derived from the human BAC library RPC1-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@jmg.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: 77
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..553

/organism="Homo sapiens"
 /db_xref="GDB:7535384"
 /db_xref="taxon:9606"
 /clone="RPC1-11-93C9"
 /clone.lib="RPC1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
 RPC111 Human Male BAC library"
 BASE COUNT 170 a 107 c 114 g 162 t
 ORIGIN

alignment_scores:

Quality: 298.00 Length: 74
 Ratio: 4.656 Gaps: 1
 Percent Similarity: 86.486 Percent Identity: 81.081

alignment_block:

US-09-697-089-2 x AQ320928/rev ..

Align seg 1/1 to reverse of: AQ320928 from: 1 to: 553

```

872 lileaspargmetasnvalleugluinleuthralaleumetleuprort 888
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
259 GTGACAGAGATGAACGTCGTAAGAAAGCTACCGCACTGATGTCGCCCTG 210
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
888 pgllyCysaspvalglnglyserleuserSerleuleulysHisleuglu 905
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
209 GGCGCTGTGAGCTGCAAGGACGCTGAGACGCCCTGTGAAACATTTGGAG 160
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
905 luvalProGlnleuvallysleuglyleulysasntrpargleuthr asp 921
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
159 AGTCCACACACTGCTCAAGCTTGCGGTGAAAACGTGAGACTCACAGAT 110
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
922 ThrGluilearglleleuglyalapheneglylysasnPro..... 935
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
109 ACAGAGATTGAAATTTAGGTAGTACACACATACAGAGCCAGATTAAT 60
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
936 .....leulysasnPhe 939
   :::::::::::::::::::: ::::::::::::::::::::::::::::::
59 GGATTTGGCCCTTAAATAATTC 38

```

seq_name: gb_gss:AQ283886

seq_documentation_block:

LOCUS AQ283886 219 bp DNA GSS 27-APR-1999
 DEFINITION RPC111-78E13.TV RPC1-11 Homo sapiens genomic clone RPC1-11-78E13,
 DNA sequence.

ACCESSION AQ283886
 VERSION AQ283886.1 GI:3910204
 KEYWORDS GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 219)
 Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
 Berry K., Granger D., Suh E., Wible C., de Jong P. and Venter J.C.
 Use of human BAC End Sequences for Sequence-Ready Map Building

TITLE

Unpublished (1998)

JOURNAL

Contact: Mark Adams

COMMENT

Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: 17
 Class: BAC ends.

FEATURES

source
 1. 219
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="GDB:752676"
 /db_xref="taxon:9606"
 /clone="RPCI-11-78E13"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
 RPCI11 Human Male BAC Library
 69 a 45 c 46 g 59 t

alignment_scores:
 Quality: 260.00 Length: 76
 Ratio: 4.194 Gaps: 1
 Percent Similarity: 81.579 Percent Identity: 72.368

alignment_block:
 US-09-697-089-2 x AQ283886 ..

Align seg 1/1 to: AQ283886 from: 1 to: 219

```

701 AlaglySerLeuValleuSerThrcysAsnileuSerle 717
|||||
3 GCTGGAACCTTACTGTCTCTCCTCAGCACCCTGTAAACATTATTCCT 52
717 UmetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThr 734
|||||
53 CATGCTGAGACCCAGTCCCTGACCATGAGAGAGAGAGACATCATCAT 102
734 erValThrAsnleuSerThrleuSerIleHisAspGluAsnGluArg 750
|||||
103 GTGTAACAACTGAAACCTTGATTCATGACCTACAGAAATCAACGG 152
751 leuProGlyGlyLeuThrAspSerLeuGlyAsnleuGlyAsnleuThrly 767
|||||
153 CTGCCGGGT.....ATTGTTAATATCAATG 178
767 sLeuIleMetAspAsnileuGlyMetAsn 776
|||||
179 TGTGTGCTTGTCTCACTAAATAAATAAT 206

```

seq_name: gb_gss:AQ112439

seq_documentation_block:

LOCUS AQ112439 630 bp DNA GSS 29-AUG-1998
 DEFINITION CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA sequence.

ACCESSION AQ112438
 VERSION AQ112439.1 GI:3484599

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 630)
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building (1998)

COMMENT

Other_GSSs: CIT-HSP-2372C1.TF
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES

source
 1. 630
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2372C1"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII"
 HindIII
 188 a 126 c 124 g 192 t

alignment_scores:
 Quality: 255.00 Length: 48
 Ratio: 5.312 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-697-089-2 x AQ112439 ..

Align seg 1/1 to: AQ112439 from: 1 to: 630

```

784 AlaglyleuLysAsnleuLysMetCysLeuPheHisleuThrH 800
|||||
486 GCTGAAGCCCTGAAACCTGAAAGAGATGTTATTATTCATTGACCA 535
800 sLeuSerAspIleGlyGlyMetAspTyrIleValIleSerleuThr 817
|||||
536 CTTCCTGACATTTGAGAGGAAATGATTACATGTCAGTCTGTCAA 585
817 erGluProCysAspLeuGluIleGlnleuValSerCysCys 831
|||||
586 GTGAACCTGTGACCTGACAAATTCATTAATTCCTCTGCTGC 629

```

seq_name: gb_est2:BF903662

seq_documentation_block:

LOCUS BF903662 251 bp mRNA EST 18-JAN-2001
 DEFINITION IL2-WT0180-181200-276-F03 WT0180 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF903662
 VERSION BF903662.1 GI:12295121

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 251)
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Nagai,M.A., da Silva,W. Jr., Zagzo,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.D., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?li=IL2&t2=IL2-MT0180-
181200-2/6-F03&t3=2000-12-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 96.
Location/Qualifiers

FEATURES
source
1..251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0180"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site.1: Sma1; Site.2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 72 a 62 c 56 g 61 t
ORIGIN

alignment_scores:
Quality: 248.50 Length: 79
Ratio: 3.883 Gaps: 3
Percent Similarity: 81.013 Percent Identity: 74.684

alignment_block:
US-09-697-089-2 x BF903662 ..

Align seg 1/1 to: BF903662 from: 1 to: 251

```

337 ThProLeupheValIleThrCysAlaIleGlnMetGly_GluSerG 353
|||||
3 ACCGCTCTTTGAGGTCATCTTGTGCATCTCATGACAGAAAGG 52
|||||
353 lUpHeHisSerHisThrGlnThrThLeupHeHisThrPheTyraSPleu 369
|||||
53 AGTTCACCTCTCACACAAACACCTGATCCATCTTATGATCTG 102
|||||
370 LeuIleGlnLysAsnLysHisLysHisLysGlyValAlaAlaSerAsp 386
|||||
103 AGATACAGAAACAAACAAACATAAAGGCGTG.....TCTGATGT 146
|||||
386 eileArGser...LeuAspHisCysGlyAspLeuAlaLeuGluGlyValP 402
|||||
147 GACTCATTCGAGCGCTGACCATCTGTGG.GACCTAAGATCTGTGGTGGCT 195
|||||
402 heSerHisLysPheAspPheGlnLeuGlnAspVal 413
|||||
196 TCTGCCACCAAGTTGAATTCGAATGCACTGACAGATGTG 230
|||||

```

seq_name: gp_gss:AZ614134

seq_documentation_block:
LOCUS AZ614134 625 bp DNA GSS 13-DEC-2000
DEFINITION IM0442N17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0442N17 R, DNA sequence.
ACCESSION AZ614134
VERSION AZ614134.1 GI:11736324
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
1 (bases 1 to 625)
Dunn,D., Aoyagi,A., Barber,M., Beccorn,T., Duval,B., Hamll,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Kelly
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0442 row: N column: 17
Seq primer: CACACAGCAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 625.
Location/Qualifiers

FEATURES
source

1..625
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g1147321141gblAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 171 a 143 c 144 g 167 t
ORIGIN

alignment_scores:
Quality: 244.00 Length: 188
Ratio: 1.952 Gaps: 4
Percent Similarity: 66.489 Percent Identity: 34.043

alignment_block:
US-09-697-089-2 x AZ614134/rev ..

Align seg 1/1 to reverse of: AZ614134 from: 1 to: 625

```

325 lIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProLeupheVa 341
|||||
617 TTGGACAGATGAAGATTTCACAGGAATTCACAAGACCCCTCTTCGT 568
|||||
341 ValIleThrCysAlaIleGlnMetGlyGluSerGluPheHisSerHisT 358
|||||
567 GGCAGCAGTATGTACTGACTGTGTTGAAAATTCATCTGACACACCCCTTC 518
|||||
358 hrcGlnThrThrLeupHeHisThrPheTyraSPleuLeuIleGlnLysAsn 374
|||||

```


THIS PAGE BLANK (uspro)

OM of US-09-697-089-2 to: GenEmbl:* out_format : pfs
Date: Mar 26, 2002 2:50 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-O=/gen2.1/USPROJ/US09697089/Unat_25032002.105137.11152/app-query.fasta.1.1097
-DB=genEmbl -QFMT=fastap -SUFFIX=ige -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.050 -LOOPEXT=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELTOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pis -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09697089.CCNI.1.17624
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-697-089-2
Query length: 1024
Database: GenEmbl:*
Database sequences: 1472140
Search length: 341344837
Search time (sec): 1827.280000

score list:

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
gb.pr:AY032589	+	5281.00	6081.17	0.0	3133	AY032589 Homo sapiens caspase
gb.pr:AY035331	+	5275.00	6074.17	0.0	3219	AY035331 Homo sapiens ICS-pro
gb.pr:AF376061	+	5271.00	6068.43	0.0	3361	AF376061 Homo sapiens caspase
gb.pr:AY027787	+	5262.00	6058.60	0.0	3396	AY027787 Homo sapiens CLANA (C
gb.hlg:AC010968	-	3733.00	4253.70	1.2e-228	160583	AC010968 Homo sapiens chrom
gb.pr:CN001053	-	3729.00	4250.60	1.7e-228	138909	AI121653 BAC sequence from t
gb.pr:IR2005817	+	2244.00	2585.57	9.6e-136	1355	AL389334 Homo sapiens mRNA fu
gb.pr:AY027788	+	1500.00	1727.29	6.1e-88	1395	AY027788 Homo sapiens CLANC (C
gb.pr:AY027789	+	819.00	947.12	1.8e-44	768	AY027789 Homo sapiens CLANC (C
gb.pat:AE6510	+	655.00	735.01	1.0e-32	6228	E23943 Excessive ovulatory an
gb.pat:AE23943	+	653.00	734.01	1.3e-32	5984	E23943 Excessive ovulatory an
gb.pat:AE4598	+	653.00	734.01	1.3e-32	5984	E23943 Excessive ovulatory an
gb.pat:AE4599	+	653.00	733.77	1.3e-32	6124	U19251 Homo sapiens neuronal a
gb.pat:HS019251	+	653.00	733.77	1.3e-32	6124	U19251 Homo sapiens neuronal a
gb.pat:AF135494	+	648.50	727.43	3.0e-32	6133	AF135494 Homo sapiens clone 13
gb.ro:AF135492	+	637.50	717.46	1.1e-31	5269	AF135492 Mus musculus clone 21
gb.ro:AF135490	+	632.50	712.77	2.0e-31	4752	AF135490 Mus musculus clone 21
gb.ro:AF102871	+	632.50	712.64	2.0e-31	4815	AF102871 Mus musculus neuronal
gb.ro:AF135489	+	630.50	708.94	3.2e-31	5497	AF135489 Mus musculus clone 9
gb.ro:AF007769	+	627.50	705.74	4.9e-31	5366	AF135491 Mus musculus clone 16
gb.pat:AE23944	+	593.50	666.49	7.5e-29	5366	E23944 Excessive ovulatory an
gb.pat:AE24990	+	593.50	666.49	7.5e-29	5366	E24990 Apoptosis-inhibiting pr
gb.pat:AE4509	+	561.50	629.30	8.8e-27	5502	E4509 Sequence 1 from Patent
gb.ro:AF242431	+	539.50	571.75	1.4e-23	117791	AF242431 Mus musculus neuron
gb.hlg:AC010272	-	539.50	568.85	2.1e-23	135306	AC010272 Homo sapiens chrom
gb.pr:AC022119	-	539.50	568.80	2.1e-23	155974	AC022119 Homo sapiens chrom
gb.pr:HS010017	+	539.50	567.97	2.1e-23	168814	AC010237 Homo sapiens chrom
gb.pr:HS010017	+	538.50	569.47	1.9e-23	131078	U08017 Homo sapiens basic tr
gb.pr:AC005031	+	538.00	573.70	1.1e-23	82876	AC005031 Homo sapiens PI clor
gb.pr:AC044797	+	538.00	564.95	3.4e-24	190871	AC044797 Homo sapiens chrom
gb.ro:AF242431	+	533.50	575.75	8.5e-24	41613	AF242433 Mus musculus delta N
gb.hlg:AC012369	-	530.00	557.85	8.4e-23	155809	AC012369 Homo sapiens chrom
gb.ro:AF131205	-	522.50	547.72	3.1e-22	179252	AF131205 Mus musculus Serf1
gb.pr:CN001058	-	497.00	519.31	1.2e-20	162692	AI121658 BAC sequence from t
gb.hlg:AC011332	-	497.00	517.92	1.4e-20	185281	AC011332 Homo sapiens chrom
gb.ro:AF242431	-	465.00	521.42	9.0e-21	90650	AF242432 Homo sapiens Naip3 g
gb.pr:AY027790	+	465.00	541.60	6.8e-22	578	AY027790 Homo sapiens CLAND (C
gb.ro:AF242433	-	333.90	2.5e-10		14784	AF242435 Mus musculus delta N

gb.pat:AX082205 + 298.50 330.92 3.7e-10 3382 AX082205 Sequence 7 from Pa
gb.pr:AF126484 + 298.50 330.30 4.0e-10 3589 AF126484 Homo sapiens CARD4
gb.pr:AF113925 + 298.50 328.10 5.2e-10 4390 AF113925 Homo sapiens Nod1
gb.sts:G55566 - 298.00 349.35 3.5e-11 553 G55566 SHGC-100923 Human Hom
gb.pat:AX082207 + 296.00 329.80 4.3e-10 2859 AX082207 Sequence 9 from Pa

seq_name: gb.pr:AY032589

seq documentation block: 3133 bp mRNA

LOCUS AY032589 Homo sapiens caspase recruitment domain protein 12 mRNA, complete

DEFINITION cds.

ACCESSION AY032589

VERSION AY032589.1 GI:13899172

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

alignment scores: Quality: 5281.00 Length: 1024
Ratio: 5.157 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-697-089-2 x AY032589 ..

Align seg 1/1 to: AY032589 from: 1 to: 3133

```

1 MetAspPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyLeu 17
|||||
36 ATGAATTTCATTAAGGACAAATGACCGAGCCCTTATTCAGAAAGATGGGAAT 85
|||||
17 ThrValIleLysGlnIleThrAspAspLeuPheValIleTrpAsnValLeuA 34
|||||
86 GACTGTTATAAAGCAATACAGATGACCTATTGTATGGAATGTTCTCA 135
|||||
34 snArgGluGluValAsnIleIleCysGlyLysValGluGlnAspAla 50
|||||
136 ATCCGCAAGAGTAACATCATTTGCTGCGAGAGGTGGACGAGATGCT 185
|||||
51 AlaArgGlyIleIleHisMetIleLeuLysGlySerGluSerCysAs 67
|||||
186 CCTAGAGGATCATTCACATGATTTTGAAGAGGTTTCAGAGTCTGTAA 235
|||||
67 IleuPheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAspL 84
|||||
236 CCTCTTCTTAAATCCCTTAAGAGAGTGAACATACCTTATTTTCAGAGCT 285
|||||
84 euAsnGlyLysIleSerLeuPheHisGlnThrSerGluLysAspLeuAsp 100
|||||
286 TGAATGACAAAGTCTTTTTCATCAGACATCAGAAAGAGACTTGGACGAT 335
|||||
101 LeuAlaGlnAspLeuLysAspLeuTrpHisThrProSerPheLeuAsnPh 117
|||||
336 TTGGCTCAGAGATTAAAGGACTTGTACATACCCCATCTTTCTGAACTT 385
|||||
117 eTrpProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerThrP 134
|||||
386 TTATCCCTTGCTGAAAGATATGACATATTTTAACTTGAAGAGCACT 435
|||||
134 heThrGluProValLeuTrpArgLysAspGlnHisHisArgValGlu 150
|||||
436 TCACAGAACTGCTGCTGAGGAGGAGACCAACCATCAGCCGCGTGAG 485
|||||
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
|||||
486 CAGCTGACCTGAAATGGCTCTGAGGCTTTCAAGAGCCCGCATCAT 535
|||||
167 eGluGlyGluSerGlyLysSerThrLeuLeuGlnArgIleAla 184
|||||
536 TGAAGGGGAAATCTGGCAAGGCAAGCTGCTGCGAGCGCATTTGCCA 585
|||||
184 eLeuTrpLysSerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
|||||
586 TGCTCGGGGCTCCGGAAGTGCAGAGCTTCGACCAAGTTCAAAATTCGTC 635
|||||
201 PhePheLeuArgLeuSerArgAlaGlnGlyLysLeuPheGluThrLeu 217
|||||
636 TTCTTCTCTCCGTCAGCAGGAGCCGAGGCTTTTGAACCCCTCTG 685
|||||
217 sAspGlnLeuLeuAspIleProGlyThrIleArgLysGlnThrPheMetA 234
|||||
686 TGATCAACTCTCGATATACCTGGCAATTCAGAGGAGACATTCATG 735
|||||
234 IAspLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTrp 250
|||||
736 CCATGCTGCTGAAGCTCGCGAGAGGCTTCTTCTTCTGATGGCTAC 785
|||||
251 AsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysG 267
|||||
786 AATGATATTCAGCCCGAGAACTGCGAGAAATCGAAGCCCTGATTAAG 835
|||||
267 uAsnHisArgPheLysAsnMetValIleValThrThrThrGluCysL 284
|||||

```

```

836 AAACACCGCCTTCAGACATGCTCATGCTGCACCACTACCTAGTGGCC 885
284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
|||||
886 TGAAGCACAATACGGCAGTGTGGTCCCTGACTGCTAGGTGGGGATATAG 935
|||||
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLe 317
|||||
936 ACAGAGACAGCCGCCAGGCTCTCATCCGAGAAAGTCTGATCAAGAGCT 985
|||||
317 uAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
|||||
986 TGCTGAAGCTTGTGCTCCAAATTCAGAAATCCAGAGTCTTGAGGAATC 1035
|||||
334 euMetLysThrProLeuPheValIleIleThrCysAlaIleGlnMetGly 350
|||||
1036 TCATGAAGACCCCTCTCTTGTGCTATCATCTTGTCATTCAGATGGGT 1085
|||||
351 GluSerGluPheHisSerHisThrGlnThrLeuPheHisThrPheTy 367
|||||
1086 GAAAGTAGTTCACATCTCACACACAAACAGCTGTCATACCTTCTTA 1135
|||||
367 rasPheLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaAla 384
|||||
1136 TGATCTGTTGATACAGAAAAACAAACAAACATAAAGGTGGCTGCAA 1185
|||||
384 eArgPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
|||||
1186 GTACACTTCATTGGAGCCTGGACACACTGTGAGACCTACTCTGGAGGCT 1235
|||||
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
|||||
1236 GTGTTCTCCACAAAGTTGATTTGCAACTGCAAGATGTCTCAGCGTGA 1285
|||||
417 ngLysAspValLeuLeuThrThrGlyLeuLeuCysLysTyrrThrAlaGln 434
|||||
1286 TGAAGATGTCTGCTGACAACTGGGCTCTCTGTAAATATTAACAGCTCAA 1335
|||||
434 rGPheLysProLysTyrrLysPhePheHisLysSerPheGlnLysTyrr 450
|||||
1336 GGTTCAGGCAAGATTAATTTCTTTCACAAGTCATTCAGGAGTACACA 1385
|||||
451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluVal 467
|||||
1386 CGAGGACGAAGACTCAGCACTTATGACGTCTCATAGACCCAGAGAGGT 1435
|||||
467 IThrLysGlyAsnGlyTyrrLeuGlnLysMetValSerIleSerAspIle 484
|||||
1436 GACCAAGGGGAATGTTACTTGCAGAAAAATGGTTTCCATTTTCGACATTA 1485
|||||
484 hrSerThrTyrrSerSerLeuLeuArgTyrrThrCysGlySerSerValGlu 500
|||||
1486 CATCCCATATATAGCAACCTCGCTCCGTTACACCTGTGGCTCATCTGGGA 1535
|||||
501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrrGlnHisGly 517
|||||
1536 GCCACAGGAGCTGTATGAAGCACCTCGACAGATGATCAACAGCGGCTG 1585
|||||
517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnLysSerL 534
|||||
1586 CCTTCTCGGACTTTCATCGCCAGAGGCTCTCTGAGACAGGAATCTT 1635
|||||
534 euGlnSerValLysAsnThrThrGluGlnGluIleLeuLysAlaIleAsn 550
|||||
1636 TGCAAAGTGTGAAGAAACACCACTGAGCAAGAAATTTGAAACCCATTAAC 1685
|||||
551 IleAsnSerPheValGluCysGlyIleHisLeuTyrrGlnLysSerThr 567
|||||
1686 ATCAATTCCTTTGTAGAGTGTGCAATCCATTTATATCAAGAGATACATC 1735
|||||
567 rLysSerAlaLeuSerGlnLysPheGluAlaPhePheGlnGlySerL 584
|||||
1736 CAATCAGCCCTGAGCCAAAGAAATTTGAAGCTTTCTTCAAGGTAAGCT 1785
|||||

```

584 euryllleasSergLYAsnIleProaspTYrIleuPheaspPheGlu 600
 1786 TATATATCAACTCAGGAGACATCCCATTTACTTATTTGACTCTTTTGA 1835
 601 HIsleuProaspncysAlaSerAlaIleuaspPheIleuLeuaspPheTY 617
 1836 CATTTGGCCCAATTTGCAAGTGTCTGGACTCTATTAACGTGACCTTTTA 1885
 617 rGlyGlyAlaMetAlaSerTPGluIuLyAlaAlaGluaspThrGlyGly 634
 1886 TGGGGGAGCTATGGCTTCATGGGAAAAGGCTGCAGAGACACAGCTGAA 1935
 634 lHIsMetGluGluAlaProGluThrTYrIleProSerArgAlaValSer 650
 1936 TCCACATGGAGAGAGCCCGCAAAACCTACATTTCCACAGGCGCTGTACT 1985
 651 LeuPhePheaspnTPlysgInguPheargThrLeuGluValThrLeuAr 667
 1986 TTGTTCTTCAACTGGAAGCAGGAATTCAGGACTCTGGAGGTCAACATCCG 2035
 667 gAsPheSerIleuAsnIlysgInaspIleThrTYrLeuGlyIlyIleP 684
 2036 GGATTTTCAGCAAGTTGAATAAGCAAGATATCATATCTGGGGAAAAATAT 2085
 684 hEsSerAlaThrSerIleuArGluGlnIleuIlyAsnIleuValGluVal 700
 2086 TCAGCTTCCACACAGCCCTCAGGCTGCANATTAAGAGATGGCTGGGTG 2135
 701 AlaGlySerLeuSerIleuValLeuSerThrCysIlyAsnIleuTYrSer 717
 2136 GCTGGAAGCCGCACTTTGGTCTCAGCACCGTGAAGAACATTTTCTCT 2185
 717 uMetValGluAlaSerProIleuThrIleGluaspGluArgHisIleThrs 734
 2186 CATGGTGAAGCCAGTCCCTCACCATAGAAGATGAGAGCACATCACAT 2235
 734 eValThrAsnIleuIlyThrLeuSerIleHisaspIleuGlnAsnGluArg 750
 2236 CTGTAACAAACCTGAAAACCTGTAGTATTATGACTACAGAAATCAACCG 2285
 751 LeuProGlyIleuThrAspSerIleuGlyAsnIleuIlyAsnIleuThr 767
 2286 CTGCGGGGTGTGACTGACAGCTGGTAACCTTGAAGAACCTTACAAA 2335
 767 sIleuIleMetAspAsnIleuIlyMetAsnGluGluAspAlaIleuIly 784
 2336 GCTCATATGATGAATAAAGATGAAGAAGATGCTATATAAACTAG 2385
 784 lAsIuGlyIleuIlyAsnIleuIlyMetCysLeuPheHisIleuThrHis 800
 2386 CTGAGGCGCTGAAAACCTGAGAGAGATGTGTTTATTTCAITTTGACCCAC 2435
 801 LeuSeraspIleGlyGluIlyMetAspTYrIleValIlySerLeuSer 817
 2436 TTGCTGTACATTTGGAGAGGAATGATTAAGTCAAGTCTGTGCAAG 2485
 817 rGluProCysaspLeuGluIuIleGlnIleuValSerCysCysLeuSer 834
 2486 TGAAGCCCTGAGCTTGAAGAAATTCATTAAGTCTCTGCTGTGCTG 2535
 834 lAsnAlaValIlyIleIleuAlaGlnAsnIleuHisaspIleuValIly 850
 2536 CAATGACAGTGAAGAAATCTAGCTCAGATCTTCAATTTGGTCAAACTG 2585
 851 SerIleuaspLeuSerGluAsnTYrIleuGluIlyAspIlyAsnGluAl 867
 2586 AGCATTTCTTGTATTCAGAAATTTACCTGGAAAAAGATGAATGAAGC 2635
 867 aLeuHisGluIleuIleuAspArgMetAsnValLeuGluGlnIleuThral 884
 2636 TCTTCATGAACCTGATCGACAGGATGAACGTGTAGAACACCTCACCCGAC 2685

884 eumelleProTPGlyCysaspValGlnGlySerLeuSerLeu 900
 2686 TGATGCTGCCCTGGGCGCTGTGACCTGCCAAGCAGCTGAGCAGCTTTG 2735
 901 LysHisLeuGluGluValProGlnIleuValIlyLeuGlyLeuIlyAsnTr 917
 2736 AAACATTTGGAGGAGGTCGCCAACACTGTCACAGTTGGGTTGAAAACTG 2785
 917 PaGluThrAspThrGluIuIleArgIleLeuGlyAlaPhePheGlyIly 934
 2786 GAGACTCACAGATACAGAGATTGAATTTAAGGTCATTTTTCGAAAGA 2835
 934 snProLeuIlyAsnPheGlnIleuAsnIleuAlaGluAsnArgValSer 950
 2836 ACCCTTGAAAAACCTTCCACAGCTTGAATTTGGCGGAAATCGTGTACC 2885
 951 SeraspIlyTrPleuAlaPheMetGlyValPheGluAsnIleuIlyGln 967
 2886 AGTGATGATGGCTTGGCTTCATGCGGTGTATTGAGAACTTTAAGCAATT 2935
 967 ValPhePheaspPheSerThrIlysgIuPheLeuProaspProAlaLeu 984
 2936 AGTGTCTTTGACTTGTAGTAAAGAAATTCACATGATCCACCATTAG 2985
 984 aAlaGlyLeuSerGlnValLeuSerIlySerLeuThrPheGlnGluAla 1000
 2986 TCAGAAAACCTTAGCCAGAGTGTATCCAGTTAACTTTTTCGCAAGAACT 3035
 1001 ArgLeuValGlyTrpGlnPheaspAspAspLeuSerValIleuArg 1017
 3036 AGGCTGTGGTGCGCAATTTGATGATGATGATGATGATGATGATGATG 3085
 1017 yAlaPheIlyLeuValThrAla 1024
 3086 TGTCTTTAACTAGTAACTGCT 3107
 seq_name: gb_pr:AY035391
 seq_documentation_block:
 LOCUS AY035391 3219 bp mRNA PRI 24-JUL-2001
 DEFINITION Homo sapiens ICE-protease activating factor mRNA, complete cds.
 ACCESSION AY035391
 VERSION AY035391.1 GI:14334214
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 (bases 1 to 3219)
 Poyet,J.L., Srinivasula,S.M., Tnani,M., Raznarz,M.,
 Fernandes-Alnemri,T. and Alnemri,E.S.
 Identification of Ipaf, a human caspase-1-activating protein
 related to Apaf-1
 J. Biol. Chem. 276 (30), 28309-28313 (2001)
 TITLE
 JOURNAL
 PUBMED 11390368
 MEDLINE 21359454
 JOURNAL
 AUTHORS
 Poyet,J.-L., Srinivasula,S.M., Fernandes-Alnemri,T. and
 Alnemri,E.S.
 2 (bases 1 to 3219)
 Direct Submission
 Submitted (16-MAY-2001) Microbiology and Immunology, Thomas
 Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107,
 USA
 FEATURES
 source
 1..3219
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="p21-22"
 145..3219
 /note="IPAF, CED4/Apaf-1 family member; caspase-associated
 recruitment domain containing protein"
 /codon_start=1
 CDS
 145..3219

/product="ICE-protease activating factor"
/protein_id="AAK59843.1"
/db_xref="GI:14334215"
translation="MNEIKDNRALIQMGMTVYKQITDDLEFVNNVLRREVNITICE
KVEDASRIAGIIHMLKKSSCNLFELSLKEMNPLRQDINGSLFPHOTSEGDIDDA
ODKLDYHPSPLNFYPLGSDIDIFNLKSTFEPEVLMRKDQHHRRERQULNGLDA
LQSPCITIEGSGSKSTLQRIAMLWGSCKALTKRFVFLSLSAQGLFETLGD
QLDIDPQIRKQFQFAMLLKRLQVLELDGYNFKFQNPCEILKELKHNKRNMY
VTTTECLRIHQGLTAEVGMTEDESAQALIREVILKEILAEILQIQSRCLNLI
MKTPLFVITCALIQMGSEFHSFTQTLPHFTFYLLQKNNKHKGYAASDEFISLH
CGDLALGVFSHKEDEFLODVSSYNEDVLLTTLGICVTLQREPKFKHKSQDET
AGRRSLTSHPEEYTKNGCYNOKMGSTSDTISYTSLLRTYCGSSVETARVMKH
LAAVYQGCILGLSIAKRLMRQESLOSVAKNTTBQELKAININSEVEGCIHLQEST
SKSALDSQEFAPFOGKSLYINSNIPLYLDFEHLDPQNASALDFIKLDEYGGAMAM
EKAEDDGIHMEAPETYPISRAVSLFPMKQEFRLTEVLRDPSFLNKQDIYILK
IFSSATSLRQIKRACVAGSLVLTSTCKNYSIAMEYKHLKTEDERHITSVNLKT
LSIHDLONRPLPGSLTSLGNLKLTKLMDNTKIMNEDAKIKLAEGLKNNKMLPHL
THSDIGEGMDYITKSLSEPCDLEFQIOLYSCLSAAYVITLQNLNLVKSITLDS
ENYLEKGNELHELIDRNLVLEQULRLMLPWGCDVGSLSLEKLEEVPLQVKKLL
KNMRLDTEIRILGAFFGNPLKNFQDLNLAGNVSDDGMLEFMGVFENKOLFDFE
STKEFLDPALVRLSQVLSKFLFLQEARLVGMQFDDDLSEVITGARKLVTAA"

SE COUNT 934 a 696 c 748 g 841 t
IGIN

Alignment scores:
Quality: 5275.00 Length: 1024
Ratio: 5.156 Gaps: 0
Percent similarity: 99.902 Percent identity: 99.902

alignment_block:
US-09-697-089-2 x AY035391 ..

Align seg 1/1 to: AY035391 from: 1 to: 3219

1 Metasphenelleusaspasensarargalaleuileglnargmetcylme 17
|||||
15 ATGAATTTCAATAAGGACATAGCCGAGCCTTATTCANAAGATGGAGAT 194
17 tthrval1lelysglnlethrasspasphevaltrpasnvalleua 34
|||||
195 GACTGTTTAAAGCAATCAGAGTACCTATTGTATGGAATGTTCTGA 244
34 snargglucluvalasnlelleecysgsluvalgluclnaspala 50
|||||
245 ATCCGGAAGAACTAAACATCATTTCTGCGAAGGTGAGCAGCATGCT 294
51 Alaargglyllellehismetlleuleuslysglyserglusercysas 67
|||||
295 GCTAGAGGATCATTCACATGATTTTGAANAAGGGTTCAGAGTCCGTGA 344
67 nleuhsenleusserleuysglutrpasntrproleupheglnaspl 84
|||||
345 CCTCTTTCTTAATCCCTTAAAGAGTGAACATATCCTCATTTCAAGACT 394
84 euaarglyglinsenleupheh1sglntrsergluclyaspleuasasp 100
|||||
395 TGAATGAGCAAAAGCTTTTTCATCAGACATCAGAAGGAGACTTGGACGAT 444
101 leu1aglnaspleuaspleu1ytrh1thrproserp1heleuasph 117
|||||
445 TTGGCTCAGGATTTAAAGACTTGTACCATACCCATCTTTTCTGAACCT 494
117 e1ytrproleuglygluasplleaspl1ellep1heasleuysserthr 134
|||||
495 TTATCCCTTGTGTAAGATATTCATATTTTAACTTGAAGAACACCT 544
134 he1hgluprovalleu1trp1arglyaspl1n1h1sh1sargvalglu 150
|||||
545 TCACAGAACCTGTCTCTGTGAGGAAGACCAACACCATCACCGCGTGAAG 594
151 glnleu1trleuasnglyleu1eugln1aleu1glnserprocy1le1 167
|||||

595 CAGCTGACCCGTGAAGGCCCTCTGACAGGCTTTCAGAGCCCTGCATCAT 644
167 egluclgluserglylysglylysserthrleu1eugln1argleu1am 184
|||||
645 TGAAGGGAATGTGCAAGGCAAGTCCCTGCTGCGACGCAATTTGCCA 694
184 e1leu1trp1yserglylyscyslysalaleu1thrlysphe1ysp1heval 200
|||||
695 TCCTCGGGCTCCGGAAAGTGCAGAGGCTCTGACCAAGTTCAATTCGTC 744
201 phe1he1leu1argleu1ser1arg1aleu1ngly1leu1p1hegl1uth1leucy 217
|||||
745 TTCTTCTCTCTGCTCAGCAGGGCCCAAGGTGACCTTTTGAAGCCCTG 794
217 saspl1leu1leuaspl1leprogl1ythr1learglysglntrp1hmeta 234
|||||
795 TGATCACTCCCTGGATATACCTGCGACATTCAGGAAGCAATTCATGG 844
234 1amelleu1leuasleu1arg1n1argval1leu1p1heleu1euaspl1ytr 250
|||||
845 CCATGCTGCTGAAGCTGCGCAGAGGCTTCTTCTCTGATGAGCTAC 894
251 asng1up1he1ysprogl1nasncysprogl1uileg1u1aleu1le1ysgl 267
|||||
895 AATGAATTCAGCCCGAGACTGCGCAGAAATCGAAGCCCTGATTAAGGA 944
267 uasn1s1argp1he1ysasn1metval1leval1thr1tr1thr1nglu1cyl 284
|||||
945 AATCCAGCCCTTCAGAAACATGATGCTCATCCACCTACACTGAGTGGC 994
284 eua1rg1s1le1eargl1n1p1hegl1y1aleu1thr1alag1uval1gl1asp1met 300
|||||
995 TGAGGCACATACGCGCAGTTGGTCCCTGACCTGACGTCAGAGTGGGATATG 1044
301 thr1gluas1p1ser1arg1n1aleu1learg1uval1leu1le1ysgl1ue 317
|||||
1045 ACAGAAAGACAGCCCGAGGCTCTCATCCGAGAGTGCATGATTAAGAGGT 1094
317 u1alegl1u1leu1leu1leu1gln1legl1n1ys1ser1arg1cys1leu1arg1asn1 334
|||||
1095 TCCTGAAGCTTGTGCTCCAAATTCAGAAATCCAGAGTCTTGAGAGAAAC 1144
334 eumelle1ys1thr1proleup1heval1le1thr1cys1ale1legl1n1metcyl 350
|||||
1145 TCATGAGAGACCCCTCTTGTGTCATCATTGTCATCCAGATGAGGT 1194
351 gluserglu1p1he1h1ser1h1thr1n1tr1p1he1h1s1thr1p1he1y 367
|||||
1195 GAAAGTGAAGTTCACCTCTCACACACAAACAGCGTGTCCATACCTCTTA 1244
367 raspleu1leu1legl1n1lys1asn1lys1h1s1lys1sgly1val1ala1as 384
|||||
1245 TGATCTGTTGATACAGAAACAAACAAACATTAAGTGTGGGTGCA 1294
384 era1s1p1he1le1ear1ys1er1leuas1p1h1s1cys1gl1yaspleu1aleu1gl1u1 400
|||||
1295 GTGACTTCATTCGGAGCCCTGAGACCACTGTGAGACCTCTGAGAGGT 1344
401 val1p1he1ser1h1s1lys1p1heas1p1hegl1u1eugln1nas1pval1ser1se1val1as 417
|||||
1345 GTGTTCTCCACAGATTTGATTTGCAACTGCAAGATGTGTCCAGGCTGA 1394
417 ngluas1pval1leu1leu1thr1thr1gly1leu1eucys1ys1ytr1p1ral1aglna 434
|||||
1395 TGAGAGTGTCTGCTGACAACTGGGCTCCTCTGTAATATATACAGTCAAA 1444
434 rgh1p1he1ys1pro1ys1ytr1lys1p1hep1heh1s1lys1ser1p1hegl1n1u1ytr1thr 450
|||||
1445 GCTTCAAGCAAGATATTAATCTTTCAACAAGTCAATTCAGAGATACACA 1494
451 ala1gl1ytr1arg1leu1ser1ser1leu1tr1ser1h1s1gl1up1rogl1u1val1 467
|||||
1495 GCAGGACGAAGACTCAGCAGTTATTTAGCTCTCATGAGCCAGAGAGGT 1544

```

467 IThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleT 484
|||||
1545 GACCAAGGGAATGTTACTTGCAGAAAATGTTCCATTCCGACATTA 1594
|||||
484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
|||||
1595 CATCCACTTAAAGCAGCTGCTCCGGTACACCTGCGGTCACTGTGTGAA 1644
|||||
501 AlaThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyC 517
|||||
1645 GCCACGAGGCTGTATGAAACACCTCGCAGCAGATGATCAACAGCGCTG 1694
|||||
517 sLeuLeuGlyLeuSerIleAlaIleAlaValArgProLeuTrpArgGlnSerL 534
|||||
1695 CCTTCCTCGACTTTCATCGCCAGAGAGGCTCTCTGGAGACAGCAATCTT 1744
|||||
534 euGlnSerValLysAsnThrThrGlnGlnGluIleLeuLysAlaIleAsn 550
|||||
1745 TGCAAAGCTGAAAAACACCACTGAGCAAGAAATTCGAAAGCCATTAAC 1794
|||||
551 IleAsnSerPheValGluCysGlyTyrIleHisLeuTyrGlnGlnSerThr 567
|||||
1795 ATCAATTCTTTTGAAGTGTGGCATCATTTATATCAAGAGGTACTATC 1844
|||||
567 rLysSerAlaLeuSerGlnGlnIlePheGluAlaPhePheGlnGlnLysSerL 584
|||||
1845 CAAATCACCCCTGAGCCCAAGAAATTTGAAGCTTTCTTTCAGAGCTAAAGCT 1894
|||||
584 euTyrIleAsnSerGlyLysAsnIleProAspTyrLeuPheAspPheGlu 600
|||||
1895 TATATATCACTCAGGGAACATCCCGATTTACTTATTTGACTCTTTGAA 1944
|||||
601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
|||||
1945 CATTTGCCCAATTTGCAAGTGCCTCGACTTCATTTAACTGGACTTTTA 1994
|||||
617 rGlyValMetLysSerTrpGluLysAlaAlaGlnAspThrGlnGlyTyr 634
|||||
1995 TGGGGAGACTTGGCTTCATGGAAAAAGCTGCCAGAACACACAGCTGCA 2044
|||||
634 LeuHisMetGluGluAlaProGluThrTyrIleProSerArgAlaValSer 650
|||||
2045 TCCACATGAGAGAGGCCCCAGAAACCTACATTTCCACAGAGCGGTATCT 2094
|||||
651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValIleThrLeu 667
|||||
2095 TTGTTCTTCACTGGAACAGCAAGAAATTCAGGACTCTGGAGGTCAACCTCG 2144
|||||
667 gAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlnLysIleP 684
|||||
2145 GGATTTTCGCAAGTTGAATTAAGCAAGATATCATGATTCGGGGAAAAATAT 2194
|||||
684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
|||||
2195 TCACCTCTGCCACAAGCTCAGGCTGCAAAATTAAGAGATGTGCTGGTGTG 2244
|||||
701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
|||||
2245 GCTGGAAACCTCAGTTGGTCTCAGCACCGTGAAGAACATTTATTTCTCT 2294
|||||
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
|||||
2295 CATGTGGAAACCACTCCCTCACCATTAAGAGATGAGAGGACATCATCAT 2344
|||||
734 euValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
|||||
2345 CTGAACAAACCTGAAAACTTGAATATTAATCAATCTACAGAAATCAACAG 2394
|||||
751 LeuProGlyGlyLeuThrAspSerLeuGlnLysLeuLysAsnLeuThrLys 767
|||||
2395 CTGCCGGGTGTGTGACTGACAGCTTGGGTAACTTGAAGAACCTTACAAA 2444
|||||

```

```

767 sLeuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeu 784
|||||
2445 GCTCATATATGATTAACATAAAGATGAATGAAGAAGATGCTTAAACTAG 2494
|||||
784 IaGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
|||||
2495 CTGAAGGCCCTGAAAAACCTGAAGAAGATGTGTATTTATTTCAATTGACCAC 2544
|||||
801 LeuSerAspIleGlyGlnGlyMetAspTyrIleValLysSerLysSerSe 817
|||||
2545 TTGTCTGCATTTGAGAGGGAATGATTAATCAATCAAGTCTGTCAAG 2594
|||||
817 rGluProCysAspLeuGlnGluIleGlnLeuValSerCysCysLeuSer 834
|||||
2595 TGAACCCGTGTACCTTGAAGAAATTCATTTGTCTCCGTGCTGTCTGTG 2644
|||||
834 IaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
|||||
2645 CAAATGCAAGTGAATAATCTTACCTCAGAAATCTTCACAATTTGTCAAACTG 2694
|||||
851 SerIleLeuAspLeuSerGluAsnTyrLeuGlnLysAspGlyAsnGlnVal 867
|||||
2695 AGCATTTCTTGAATTTATCAGAAATTTACCTGGAAAAAGATGCAAAATGAAG 2744
|||||
867 aLeuHisGluLeuLeuIleAspArgMetAsnValLeuGlnGlnLeuThrAla 884
|||||
2745 TCTTCATGCACTGATCCAGACGATGAACGTCTTGAAACAGCTCACCCGAC 2794
|||||
884 euMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
|||||
2795 TGATGTGCTGCCCTGGCGCTGTGACGTGCAGAGCAGCTGAGAGGCTGTG 2844
|||||
901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlnLysAsnTr 917
|||||
2845 AAACATTTGGAGGAGTCCCAACACTGCTCAACCTGTGGGTGAAAACTG 2894
|||||
917 PArgLeuThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLys 934
|||||
2895 GAGACTCAGACATTCAGAGATTAAGATTTTAGTGCAATTTTGTGAAAGA 2944
|||||
934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
|||||
2945 ACCCTCTGAAAAACCTCCAGCAGTTGAATTTGGCGGAAAACTGCTGAGC 2994
|||||
951 SerAspGlyTrpLeuAlaPheMetGlyValPheGlnAsnLeuLysGlnLe 967
|||||
2995 AGTATGGAATGGCTTGCCTTCATGGGTATTTGAGAATCTTTAAGCAATT 3044
|||||
967 uValPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeu 984
|||||
3045 AGTGTTTTTCACCTTTAGTACTAAAGAATTTCTACCTGATCCAGCATTAG 3094
|||||
984 aLArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
|||||
3095 TCAGAAAACTTAGCCCAAGTGTATCCAAAGTTACTTTCTGCAAGAGCT 3144
|||||
1001 ArgLeuValGlyTrpGlnPheAspAspAspAspLeuSerValIleThrG 1017
|||||
3145 AGGCTTGTGGTGGCAATTTGATGATGATGATCTCAGTGTATTAACAG 3194
|||||
1017 yAlaPheLysLeuValThrAla 1024
|||||
3195 TGCCTTTAACTAGTAACGTCT 3216
|||||

```

seq_name: gb_pr:AF376061

seq_documentation_block: 3581 bp mRNA

LOCUS AF376061 3581 bp mRNA 15-MAY-2001

DEFINITION Homo sapiens caspase recruitment domain protein 12 mRNA, complete

ACCESSION AF376061
 VERSION AF376061.1 GI:14040074

KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3581)
AUTHORS Gingras, M., Qiu, J., and Margolin, J.F.
TITLE Differential expression of the caspase recruitment domain protein 12 (CARD12) during monocytic differentiation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3581)
AUTHORS Gingras, M., Qiu, J., and Margolin, J.F.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center, Baylor College of Medicine, 6621 Fannin St., MC3-3320, Houston, TX 77030, USA

FEATURES
SOURCE location/Qualifiers
1..3581 /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="3p21-22"
/cell_line="0937"
/cell_type="peripheral blood-derived CD14 mature monocyte"
490..3564 /note="CARD12"
/product="caspase recruitment domain protein 12"
/protein_id="AAK53443.1"
/db_xref="GI:14040075"
/translation="MNFIKDSRALIORMKNTVIKQTTDDLEFVNVNLRNEVNIICE
KVEDDAARGIIHMLIKGSESCNLFSLKEMWPLFODLNGSLFQTSGLDLDLA
QDLKDLHYTFNLFYPLGEDIDIFNLKSTFPLVMRKQDHHREYDLELNLDA
LOSCEIEGSGKSTYLORIAMGSKCALTKRFVFLSLRAOGGLFETLDD
OLDIDPGTIRKOTFMAMLIKLRORVFLDNGEFGKONCEIFALIKENRFENMVI
VTPTTECRHTRGOCALFAYGMDTSDSACALILEVILKELASLLQIQKRLRL
MKTPLEVYTICAIOMGESERHSHTOTTLFTFTDGLQKNHKKGYASDFISLDH
CGDLLEGVFSHKDFELQDVSVDALLTGTGLCKYTAQRFPKYKFEKFSQEY
AGRLSLSLTSHPEEYTKNGVYLQKNVSLSDISTYSLRLYCGSSVEATRAVMEH
LAAVYOHGCLGLSILAKRPLROESLOSVKNTTQELIKAINISFVCGIHLTQEST
SKALSOEFAEFGQKSLYINGNIPLYLPEFPHLPNCASALDFIKLDFGGMASW
EKAEADTGGIIMEAPETVYIPSRVSLFENMKOFERFLVLEPFLKLNKQIDRYLCK
IFSSATSLRLQIKKACVAGSLSLVSTCKNTYSLMEASPLTIEDERHITTSVNLKT
LSTHDNLORLPGLTDSLGNLKNLTKLMDNIMNEEDAKIKLEGLKNLKKWLFH
TLHSDIGEMDYIVKSLSEPCDLEIQLVSCLSANAVKILAEGLNKLSTLDS
ENYLEKGDNEALHELIDRMNVEJOLITLMLPGCDVGSSLSLKHLEEVQVLYKGL
KNMRLDTREIRILGAFGKNPLKNFOOLNLAGNVSSDGMLEAFVPEENKOLVFPDF
STKEFLPDPAVLRKLSVLSKTLFLQEARLVGMQFDDDLISVITGARKLVTA"

misc_feature
502..741 /note="Region: caspase recruitment domain"

BASE COUNT 1033 a 781 c 843 g 924 t

alignment_scores:
Quality: 5271.00 Length: 1024
Ratio: 5.152 Gaps: 0
Percent Similarity: 99.902 Percent Identity: 99.805

US-09-697-089-2 x AF376061 ..

Align seg 1/1 to: AF376061 from: 1 to: 3581

1 MetAsnPhelIleLysAspAsnSerArGAlaLeuIleGlnArgMetGlyLeu 17
|||||
490 ATGAAATTTTATAAAGGACAAATAGCCGACCTTATTTCAAAGATGGGAAAT 539
17 tThrValIleLysGlnIleThrAspAspLeuPheValIleTrpAsnValLeu 34
|||||
540 GACTGTATATAAAGCAATCACAGATGACCTATTGTATGGAATGTTCTGA 589
34 snArgGluGluValAsnIleIleLysCysGluLysValGluGlnAspAla 50
|||||

590 ATGCGGAAAGAAAGTAAACATCATTTGCTGCGAGAGAGGTGAGCAGATGCT 639
51 AlaArgGlyIleIleHisMetIleLeuLysLysGlySerGluSerCysAs 67
|||||
640 GGTAGAGGATCATTCACATGATTTTGAAGAAAGGGTTCAGATCCTCTGA 689
67 nLeuPheLeuLysSerLeuLysGluIleTrpAsnIleProLeuPheGlnAsp 84
|||||
690 CCTCTTCTTAATCCCTTAAGAGAGTGAACATATCCTCATTTTCAGACT 739
84 euAsnGlyLysIleSerLeuPheHisGlnThrSerGluGlyAspLeuAsp 100
|||||
740 TGAATGACAAAGCTCTTTTCATCAGACATCAGAAAGAACTTGGACCAT 789
101 LeuAlaGlnAspLeuLysAspLeuIleThrProSerPheLeuAsnPh 117
|||||
790 TTGGCTCAGGATTTAAAGGACTTGTACATACCCATCTTTCTGAACT 839
117 eTyrrProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerThr 134
|||||
840 TTATCCCTTGTGTAAGATATTGACATTATTTTAACTTGAAGACACCT 889
134 heThrGluProValLeuIleThrArgLysAspGlnHisHisArgValGlu 150
|||||
890 TCACAGAAACAGTCTCTGTGAGGAAGACCAACACCATCACCCTGGAG 939
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
|||||
940 CAGCTGACCTCAATGCTCTCTGACAGGCTCTTCAGAGCCCTGCATCAT 989
167 eGluGlyGluSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAla 184
|||||
990 TCAAGAGGAATGTGGCAAGCAAGCACTGCTGTCGACGAGCCATTTGCA 1039
184 eLeuThrProLysSerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
|||||
1040 TCCTCTGGGGCTCCGAAAGTCCAAAGCTCTGACCAAGTCCAAATTCGTC 1089
201 PhePheLeuArgLeuSerArgAlaGlnGlyLysLeuPheGluThrLeu 217
|||||
1090 TTCTTCTCTCGTCTCAGCAGGCGCCAGGGTGGACTTTTGAACCTCTCG 1139
217 sAspGlnLeuLeuAspIleProGlyThrIleArgLysGlnThrPheMet 234
|||||
1140 TCATCAACTCTCGATATACCTGCGACAAATCAGAAAGCAGACATTCATG 1189
234 IamLeuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspGlyTy 250
|||||
1190 CCATGCTGCTGAAGCTGCGGACAGAGGTTCTTCTCTTGATGAGTAC 1239
251 AsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysG 267
|||||
1240 AATGAATTCAGCCGCCAGAACCTGCCAGAAATCGAAGCCCTATTAAGGA 1289
267 uAsnHisArgPheLysAsnMetValIleValIleThrThrThrGluCys 284
|||||
1290 AAACACACCGCTTCAAGAACATGTCATGTCACACATCACATGAGTGCC 1339
284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
|||||
1340 TGAAGCACATACGCGCTTGGTGGCTGATGCTGAGGTGGGGAATTAAG 1389
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGlu 317
|||||
1390 ACAGAAAGAGAGCGCCAGGCTCTCATCCAGAAAGTGTGATCAAGAGCT 1439
317 uAlaGluGlyLeuLeuGlnIleGlnLysSerArgCysLeuArgAsn 334
|||||
1440 TCGTGAAGCTGTGTGCTCAATTCAGAAATCCAGGTCTTGAAGAAATC 1489
334 euMetLysThrProLeuPheValIleThrCysAlaIleGlnMetGly 350
|||||
1490 TCATGAAGACCCCTCTTGTGTGATCATCATTTGTGCAATTCAGATGGGT 1539

351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
1540 GAAAGTAGTCCACCTCTCACACAAACAACGCTGTTCCATACCTTCTA 1589
367 rAsPLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaIAs 384
1590 TGAATCTGTGATACAGAAAAACAACAACATAAAGGTGTGGCTGCA 1539
384 eArSPheIleArGSerLeuAspHisCysGlyAspLeuAlaLeuGly 400
1640 GTACTTCATTCGAGCCTGACACACTGTGAGACCTACCTCTGAGAGGT 1689
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
1690 GTGTGTCCTCCACAAGTTTGATTTGCAACTCAGAGATGTGTCCAGCTGAA 1739
417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyThrAlaGln 434
1740 TGAAGATGCCCTGCTGACACTGGGCTCCTGTAATATATACAGCTCAA 1789
434 rGpHeLysProLysTyThrLysPhePheHisLysSerPheGlnGlyThr 450
1790 GGTTCAGCCCAAGTAAATCTTTCACAAGTATCCAGAGTACACA 1839
451 AlaGlyValArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
1840 GCAGGACGAAGACCTACAGCTATTGACCTGTATAGCCACAGAGGAGGT 1889
467 IThrLysGlyAsnGlyTyThrLeuGlnLysMetValSerIleSerAspIleT 484
1890 GACCAAGGGAATGGTTACTTGCAAGAAATGGTTCCATTTGCGACACTTA 1939
484 hSerThrTyThrSerSerLeuLeuArgTyThrCysGlySerSerValGlu 500
1940 CATCCACTTATACAGCCTGCTCCGGTACACCTGTGGTCACTGTGGA 1989
501 AlaThrArgAlaValMetLysHisLysAlaAlaValTyGlnHisGlyCys 517
1990 GCCACCGGGCTGTATGAAGCACCTGCGACAGTGTATCAACACGGCTG 2039
517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuTyParGlnGlnSerL 534
2040 CCTTCGCACTTTCATCGCCAAAGAGCCTCTCTGAGACAGAAATCTT 2089
534 euGlnSerValLysAsnThrThrGlnGlnGlnLysLysAlaIleAsn 550
2090 TGCAAAGTGGAAAAACACACACGAGCAAGAAATTCGAAAGCCATPAAAC 2139
551 IleAsnSerPheValGluCysGlyIleHisLeuTyrcGlnGlnSerThrse 567
2140 ATCAATTCCTTTAGTAGTGTGGCATCCATTTTATCAAGAGAGTACATC 2189
567 rLysSerAlaLeuSerGlnGlnPheGlnAlaPhePheGlnGlyLysSerL 584
2190 CAATATGAGCCCTGAGCCCAAGAATTTGAAGCTTCTTCAAGGTAAAAAGCT 2239
584 euTyrrIleAsnSerGlyAsnIleProAspTyrrLeuPheAspPheGln 600
2240 TATATATCAACTCAGGGAACATCCCGATTACTTATTTGACTCTTTGAA 2289
601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
2290 CATTTGGCCCAATGTGCAGAGTGGCCTGACTTCATTTAACTGGACCTTTTA 2339
617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrGlyGlyT 634
2340 TGGGGGAGCTATGAGCTCATGGGAAAAAGCTCGCAAGAACACAGGTGCA 2389
634 IeHisMetGluGluAlaProGlnThrTyrrIleProSerArgAlaValSer 650
2390 TCACATGGAAGAGGCCCAAGAAACCTACATTCGCCACAGGCTGTATCT 2439

651 LeuPheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeuAr 667
2440 TTGTGTCCTTCACTGGAAGCAGGAATTCAGACTCTGAGAGTCACTCCG 2489
667 gaSPheSerLysLeuAsnLysGlnAspIleThrTyrrLeuGlyLysIleP 684
2490 GCATTTCCAGACAGTTGAATAGCAAGATATCAGATATCTGGGAAAAATAT 2539
684 hSeSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
2540 TCAGCTCTGCCAACAGCCTCAGCCTCAATAAAGAGATGTGCTGTGTG 2589
701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrrSerLe 717
2590 GCTGGAGGCTCAGTTTGGTCTCAGACCTGTGAACCTTAAGAACTTTATCTCT 2639
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
2640 CATGGTGGAAAGCCAGTCCCTCAACCATAGAGATGAGAGCACATCACT 2689
734 eValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
2690 CTGTAAACAAACCTGAAGAACTTGATATTCATGACCTACAGAACTCAACGG 2739
751 LeuProGlyValLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrly 767
2740 CTGCGGGGTGTCTGACTGACACTTGAGGTAACCTTAAGAACTTCAAA 2789
767 sLeuIleMetAspAsnIleLysMetAsnGlnGluAspAlaIleLysLeu 784
2790 GCTCATATAGTATACATATAAGATGAATGAAGAATGCATATAAACTAG 2839
784 IaGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
2840 CTGAGGCGCCGAAAAACCTGAAGAAGATGTGTTATTTCATTGACCCAC 2889
801 LeuSerAspIleGlyGlnGlyMetAspTyrrIleValLysSerLeuSerse 817
2890 TTGTCTGACATTGAGAGGAAATGGAATGATACATAGTCAAGTCTGTGCAAG 2939
817 rGluProCysAspLeuGlnGluIleGlnLeuValSerCysCysLeuSera 834
2940 TGAACCTGTGACCTTGAGAAATTCATATAGTCTGCTGCTGTCTG 2989
834 IaAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
2990 CAATGCAGTGAAGAAATCTTACAGATCTTCACAATTTGGTCAAACTG 3039
851 SerIleLeuAspLeuSerGluAsnTyrrLeuGluLysAspGlyLysnGluAl 867
3040 AGCATTCCTTGATTTATCAGAAAATTCCTGGAAGAAAGATGGAATAGAC 3089
867 aLeuHisGluLeuIleAspArgMetAsnValLeuGlnGlnLeuThrAlaL 884
3090 TCTTCATGAACGTATGACAGATGAAAGTGAACGTGAAGACGTACCGCAC 3139
884 euMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuLeu 900
3140 TGATGTGTGCTGCTGAGGTGTGAGGTGCAAGGCAAGCCTGAGCACCTGTGTG 3189
901 LysHisLeuGlnGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
3190 AAACATTTGGAGAGAGTCCACACACTGCTCAACCTTGCGTTGAAAACCTG 3239
917 ParGluThrAspThrGluIleArgIleLeuGlyAlaPhePheGlyLysA 934
3240 GAGACTCACAGATACAGAGATTTGAATTTAGTGATTTTGGAAAGA 3289
934 snProLeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
3290 ACCCTCTGAAAACTTCCAGCACTGTAATTTGGCGGGAATCTGTGAGC 3339
951 SerAspGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967

```

|||||
3340 AGTGTGATGCGCTTGCCTTCATGGGCTATTGTGAGAACTTAAACCAAT 3389
967 uValPhepHeaspHeSerThrLySGluPheLeuProaspProAlaLeuV 984
3390 AGTGTTTTGTACTTACTACTAAAGAAATTTCTACTGATCCAGCATTTAG 3439
984 aAlargLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
3440 TCAGAAACTTAGCCAGAGTGTATCCAGTTACTTCTGCAAGAGCT 3489
1001 ArgLeuValGlyTrpGlnPheaspaspaspLeuSerValIleThrGlu 1017
3490 AGGCTGTGGTGCGCAATTCATGATGATGATCTCAGCTTATTATACAG 3539
1017 yAlaPheLysLeuValThrAla 1024
3540 TGCCTTTAACTAGTACTGCT 3561

seq_name: gb_pr:AY027787

documentation_block:
DEFINITION Homo sapiens CLANA (CLAN1) mRNA, complete cds.
ACCESSION AY027787
VERSION AY027787.1 GI:14324112
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3396)
Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL
Genomics 75 (1-3), 77-83 (2001)
MEDLINE
21365712
PUBMED
11472070
REFERENCE
2 (bases 1 to 3396)
Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
AUTHORS
Direct Submission
JOURNAL
Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source
1. .3396
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/lssue_type="lung"
1. .3396
/gene="CLAN1"
277. .3351
/gene="CLAN1"
/codon_start=1
/product="CLANA"
/protein_id="AAK14776.1"
/db_xref="GI:14324113"
/translational="NMFKDMSRALIORMGMTVYIKOITDDLPVNVNLRREVNITCE
KVBODARGIHHILTKGSSCSNLFLKSKENNYPLFODINOSLTHQTSGLDOLA
ODKLDIHTPSFLNTPLEGDIDIIINLKSTFEPIILMKRDQHHHVEQLITGLLOA
LOSPLIIEBESGKSTLQRIAMLGSGKALTKFEVFLRLSRAGGLEETLCD
QLDIDPGRKQTFMAMLLKROVLFLDGVNEFPONCPETELAIKENHFKNNVY
VTTECLRHIRGFALTAEGMDTSDSOALIREVILKELAGLILLOKSRCLNL
MTPPLFVITCAIOMGESSEPHSHOTTFTEFYDLILOKKHKKHKGVAASDFTISDH
RGDLAEGVSHKDFELEDVSVNEDVLITLLCKYRAQRKPKYKFKHKSFOEY
AGRLSSLLSHEPEEVTGNGYLQKMSISDITSYSSLRTYSGSSVEATRAVMKH
LAAVYHGCLLGIISIAKRPIMROESLOSVKNTTEOILKAININSVEGILHYOEST
SKSALOEFEEAFQSGSLYINSGLIDYLFDEFEHLPCNASALDFIKLDFYGGASW
EKAAEDTGGIIMEADEPTYIPSAVSLFEMMKOEFPLLEVTLDFSKLNKODIRYLGK
IFSSATSLRLOIRKRCAGVAGSLIVSTCKNYSILVAASPLIIEDEHRTSTYNTKT
LSHDLOQRRLPGGLTDSIGNLKNLTKLIMDNIRKMBEDAIIKLABGILNKKKCLHL
THLSDIGEGMDYLVKSLSEPCDLEIEIQLVSCCLSANAVKIIAQNHLNVLKSLIDS

```

```

ENYLEKDGNEALHELIDRNMVLEQLTALMLPMGCDVQSSLSLKLHLEVPOLVKLGL
KNMRLDTEIRILGAFFGNKPLKNFQOLNLAGRVSSDGLMFMGVFNKOLVFEDF
STKEFLPDPALVVKRLISOVLSKTFLEQEARLVGMQFDDDLISVITGAFKLVTAA"
BASE COUNT 992 a 737 c 793 g 874 t
ORIGIN
alignment_scores:
Quality: 5262.00 Length: 1024
Ratio: 5.149 Gaps: 0
Percent Similarity: 99.805 Percent Identity: 99.707
alignment_block:
US-09-697-089-2 x AY027787
Align seg 1/1 to: AY027787 from: 1 to: 3396
1 MetAspPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyLeu 17
|||||
277 ATGAATTCATTAAGACATAAGCCAGCCCTTATTCAAAGATGGGAAAT 326
17 ThrValIleLysGlnIleThrAspAspLeuPheValTrpAsnValLeuA 34
|||||
327 GACTCTTATTAAGCAATACACGATGACCTATTGTATGGAATGTTCTGA 376
34 snArgGluGluValAsnIleIleCysCysGluIuValGluGlnAspAla 50
|||||
377 ATCCGGAAGAAGTAAACATCATTTCTCGAGAAGGTGAGAGAGATGCT 426
51 AlaArgGlyIleIleIleHisMetIleLeuLysGlySerGluSerCysAs 67
|||||
427 GCTAGAGGATCATTCACATGATTTTGAAGAGGGTTCAGAGTCTGTAA 476
67 nLeuPheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAspL 84
|||||
477 CCTCTTTTAAATCCCTTAAAGAGTGAACCTATCTTATTTCAAGACT 526
84 euAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
|||||
527 TGAATGACAAAGTCTTTTTCATCAGACATCAGAAGGAGACTGGAGCAT 576
101 LeuAlaGlnAspLeuLysAspLeuThrHisThrProSerPheLeuAsnph 117
|||||
577 TTGGCTCAGATTAAAGACTTTCACATACCCCATCTTTCTTGCAGACTT 626
117 eTyPProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerThrP 134
|||||
627 TTATCCCTTGGTGAAGATATTGACATTATTTTACTTGAAGACACT 676
134 heThrGluProValLeuThrParGlyAspGlnHisHisArgValGlu 150
|||||
677 TCACAGAACTATCTCTGTGAGGAAGCAACACCATCACCAGGTGAG 726
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
|||||
727 CAGCTGACCTGAAATGGCTCTCTGAGGCTCTTCAGACCCCTGCATCAT 776
167 eGluGlyGlnSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAla 184
|||||
777 TCAAGGGAATCTGGCAAGGCAAGCTCACCTCTGTCGAGGCAATGGCA 826
184 etLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
|||||
827 TGCCTGGGGCTCCGGAAGTGAAGGCTCGACCAACTTCAAAATTCCTC 876
201 PhePheLeuArgLeuSerArgAlaGlnGlyGlyLeuPheGluThrLeu 217
|||||
877 TTCTTCCCTCCGCTCAGCAGGAGGAGGCTGACTTTTGAAGACCTCTG 926
217 sasPheLeuLeuAspIleProGlyThrIleArgGlyGlnThrPheMet 234
|||||
927 TGATCAACTCTCGATTAACCTGCGACATCAGAGGAGGACATTCATGAG 976

```



```

234 IamLeuLeuLysLeuArgLInArgValLeuPheLeuLeuAspGlyTyr 250
|||||
977 CCATGCTCTGTAAGCTGGCGGAGGGTCTTTCTTCTTCTGATGGCTAC 1026
|||||
251 AsnLInPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysGlu 267
|||||
1027 AATGAATTCAAGCCCGCAAGACTGCCAGAAATCGAAGCCCTGATAAAGA 1076
|||||
267 uAsnHisArpPheLysAsnMetValIleValThrThrThrGluCysL 284
|||||
1077 AAACCAAGCGCTTCAAGACATGCTCATCTCACCACCTACCTAGTGC 1126
|||||
284 euArgHisIleArgGlnPheGlyAlaLeuThrAlaGluValGlyAspMet 300
|||||
1127 TGAGGCAATACGGCAGATTGGTGCCTGACTGCTGAGGTGGGATATG 1176
|||||
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLe 317
|||||
1177 ACAGAAACACAGCCCGCAGGCTCTCATCCGAGAACTGTCATCAAGAGAGCT 1226
|||||
317 uAlaGluGlyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
|||||
1227 TGTGAAAGGCTGTGCTCCAAATTCAGAAATCCAGGTGCTTGAGGATC 1276
|||||
334 euMetLysThrProLeuPheValValIleThrCysAlaIleGlnMetGly 350
|||||
1277 TCATGAAGACCCCTCTTGTGTGTCATCACTTGTGCAATCCAAATGGGT 1326
|||||
351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTy 367
|||||
1327 GAAAGTAGTCCACTCTCACACAAACAAACGCTGTTCATACCTCTTA 1376
|||||
367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaIle 384
|||||
1377 TGATCTGTTGATACGAAACACAAACAAACATAAAGGTGGCTGCA 1426
|||||
384 eArAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluGly 400
|||||
1427 GTGACTTCATTCGAGCGCTGACACCGTGGAGACCTAGCTTGAGAGGT 1476
|||||
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
|||||
1477 GTGTTCTCCACAAAGTTGATTCGAACTCGAGATGTGTCCACCGTGA 1526
|||||
417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysThrThrAlaGln 434
|||||
1527 TGAGGATGTCCTGCTGACAACTGGGCTCTCTGTAATATATACAGCTAAA 1576
|||||
434 rGpPheLysProLysTyrLysPhePheHisLysSerPheGlnGluTyrThr 450
|||||
1577 GGTTCAGACCCAAAGTATAAATTTCTTCACAAATCATTCACAGGATACACA 1626
|||||
451 AlAGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
|||||
1627 GCAGAGACGAAGACTCACAGTTTATGACGTCATGAGCCAGAGGAGGT 1676
|||||
467 LThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleT 484
|||||
1677 GACCAAGGGAGATGTTACTTGCGAATAAGGTTTCCATTTGGACATTA 1726
|||||
484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
|||||
1727 CATCACTATATAGCAGCCCTGCTCGGTACACCTGTGGTCACTGCTGGA 1776
|||||
501 AlArThrArgAlaValMetLysHisLeuAlaAlaValTyrGlnHisGlyC 517
|||||
1777 GCCACAGGGGCTGTATGAAGCACCTGCGACATGTATCAACACGGCTG 1826
|||||
517 sleuLeuGlyLeuSerIleAlaLysArgProLeuTyrArgGlnGluSerL 534
|||||
1827 CTTTCTGGACTTTCATCCGCAAGAGGCTCTTGTGAGACAGGAATCTT 1876
|||||
534 euGlnSerValLysAsnThrThrGlnGlnIleLeuLysAlaIleAsn 550
|||||
|||||
1877 TGCAAGCTGTGAAAAACACCACTGACAGAAATTTGAAAGCATTAAC 1926
|||||
551 IleAsnSerPheValGluCysGlyIleHisLeuTyrGlnLysThrSe 567
|||||
1927 ATCAATTCCTTTGTAGAGTGTGGCATCTTATATCAAGAGATACATC 1976
|||||
567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnGlyLysSerL 584
|||||
1977 CAATACAGCCCTGAGCCAGAAATTTGAAAGCTTCTTTCAGAGTAAAGCT 2026
|||||
584 euTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPhePheGlu 600
|||||
2027 TATATATCACTCAAGGAGACATCCCGATTACTTATTTGACTTCTTGAA 2076
|||||
601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
|||||
2077 CATTTGCCCAATTGTGCAGCTGCGCTGACTTCATTAACGTGACATTGTA 2126
|||||
617 rGlyGlyAlaMetAlaSerTrpGluLysAlaIleGluAspThrGlyGly 634
|||||
2127 TGGGGAGCTATGGCTTCATGGGAAAGGCTGCGAGAGACACAGTGGAA 2176
|||||
634 leHisMetGluGluAlaProGluThrTyrIleProSerArgAlaValSer 650
|||||
2177 TCACATGGAAGAGGCCCCCAAAACCTACATTCACAGACAGGCTGTATCT 2226
|||||
651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeuAr 667
|||||
2227 TTGTTCTTCAACTGGAAGCGAATTCAGAGACTGTGAGGTGCACACTCCG 2276
|||||
667 gAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
|||||
2277 GCATTTCCAGAGTTGATATAGCAAGATATACAGATATCTGGGAAATAT 2326
|||||
684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
|||||
2327 TCAGCTCTGCACACAGGCTCAGGCTGCATAATAAGATGTGCTGGTGTG 2376
|||||
701 AlAGlySerLeuSerLeuValIleSerThrCysLysAsnIleTyrSerLe 717
|||||
2377 GCTGGAAGCCTCAGTTGGTCTCAGACCTGTAAAGAACCTTATATCTCT 2426
|||||
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
|||||
2427 CATGTGGAAGCCAGTCCCTCACCATAGAAATGAGAGGCAATACAT 2476
|||||
734 eValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
|||||
2477 CTGTAACAAACCTGAAACCTTGAGTATTCATGACCTACAGATCAACGG 2526
|||||
751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
|||||
2527 CTGCGGGGTGCTGACTGCACACTTGGGTAACTTAAGAACCTTACAA 2576
|||||
767 sleuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeu 784
|||||
2577 GCTCATATGATGATATTAAGATGATGAGAAAGATGCTATAAACTGAA 2626
|||||
784 laGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
|||||
2627 CTGAAGGCTCGAAACCTGAAAGAGATGTATTATTTCAATTGACCCAC 2676
|||||
801 LeuSerAspIleGlyGluGlyMetAspTyrIleValLysSerLeuSerSe 817
|||||
2677 TTGTCTGACATTTGGAAGGAATGATTAATAGTAAAGCTCTGTCAAG 2726
|||||
817 rGluProCysAspLeuGluGluIleGlnLeuValIserCysLysLeuSera 834
|||||
2727 TGAACCTGTGACCTTGAAAGAAATTCATATTAAGTCTCTGCTGTCTG 2776
|||||
834 laAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 850
|||||

```


misc_feature 91599..116835
/note="assembly_name:Config20
clone_end:sp6
vector_side:right"
misc_feature 116936..160583
/note="assembly_name:Config21"
BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others
ORIGIN

alignment_scores:
Quality: 3733.00 Length: 1067
Ratio: 4.854 Gaps: 4
Percent Similarity: 72.071 Percent Identity: 70.947

alignment_block:

S-09-697-089-2 x AC010968 ..

align seg 1/1 to: AC010968 from: 1 to: 160583

1 MetAspHellelysAspAsnSerArgAlaLeuIleGlnArgMetGlyme 17
:::|||||
48109 GTGAAATTTCTAAAGACAAATACCGCCCTTATTCAAAGAAATGGGAAT 48158
17 tTnrvAlIleLysGlnIleThraspAspLeuPheValTrpAsnValLeuA 34
|||||
48159 GACTGTTATAAGCAATTCACAGATGACCTATTGTATGGAATGTTCTGA 48208
34 snArgGluGluValAsnIleIleCysCysGluLysValGluGlnAspAla 50
|||||
48209 ATGCGGAGAGAGTAAACATCATTTGCTGCGAGAGGAGCAGAGATGCT 48258
51 AlarArgGlyIleIleHisMetIleLeuLysGlySerGlySerCysAs 67
|||||
48259 GCTAGAGGATCATTCACATGATTTTGAAAGAGGTTCCAGATCCGTAA 48308
67 nLeuPheLeuLysSerLeuLysGluTrpAsnTyProLeuPheGlnAspL 84
|||||
48309 CCCTTTCTTAAATCCCTTAAAGAGGTGAACATACTCTATTTCCAGACT 48358
84 euAsnGlyGln..... 87
|||||
48359 TGAATGACACAAGTAAGTATCAGTTAGTTTGCATCTGTTCTCAAGTA 48408
87 87
48409 GAATAGACACAACAAGATCATTTCCCATGGCAGAGCCTCTGCTTT 48458
87 87
48459 TCCTTCTTCATATACACTGCGCAGTGGCCTTGATACCTGTAAGAAA 48508
87 87
48509 AGCTAGATGATTTTAGAAAACTCATAGATCAATTGATGATGAAG 48558
87 87
48559 CCAAAACAATAACCCCAAGGCCAGAAATAAATGATTGAGTATTGTCTCA 48608
87 87
48609 CTGCAAGATAAAACTGGAGATAAGTCACTAGTTAAGCTTCAGAGTCAG 48658
87 87
48659 CAGTGTCTTCAGCATCTCTGGTGGATGAGATCTACTATGAGAGAGA 48708
87 87
48709 GGGCCAGCTGACTACTCTCATCTGCAGGCTCTGTTAAAGACAGCC 48758
87 87

48759 TGACTGACAGACACCATGATTCATGTTTCATAAATATTGTTTTATT 48808
87 87
48809 TTTTAAATAGTAATAAAGTCAATGACAGATTTTACTCAGATGAAAGGCC 48858
87 87
48859 TACTATGCTGACAATTATCCCTAACCATCTTCTGCGCAATTCCTTAC 48908
87 87
48909 TTTCTTAACCAAAAAACAGCTCTCTCCCTTTTCCAAACGACGTGTATAA 48958
87 87
48959 GATTTGGTGTGGAAAGAGAGAAACAGAAAGCAAAAGGGAGTGGGTA 49008
87 87
49009 AAGGATGCTGGGAAAGGGGAGCAGACAGAACCTTTGTGCCAGCTCAG 49058
87 87
49059 AACTGGGAAGGACATTACATTCACCAATGGCAGATTCTCAGCCTCAGCC 49108
87 87
49109 CCCCCTTCATCTCTGAACTGAGTTTGTAGAGAGACAGAAATGGGT 49158
88 Ser. LeuPheHisGlnThrSerG 95
49159 CTTCCTTAACATCACCTCTTCTTAATAGTCTTTTTCACACACATCAG 49208
95 luGlyAspLeuAspAspLeuAlaGlnAspLeuLysAspLeuTyHisThr 111
|||||
49209 AAGGAGCTTGAGAGATTTGGCTCAGAGATTTAAAGCATGTCATCAGTAC 49258
112 ProSerPheLeuAsnPheTyProLeuGlyGluAspIleAspIleIlePh 128
|||||
49259 CCATCTTTCTGAACTTTTATCCCTTGTGGAGATATTCATTTATTTT 49308
128 euAsnLeuLysSerThrPheThrGluProValLeuTrpArgLysAspGlnH 145
|||||
49309 TAACCTGAAAGACCTTCACAGAACCTGCTGTGGAGAGAGACATAC 49358
145 IShHisArgValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeu 161
|||||
49359 ACCATCACCGCGTGAGACACTGACCTGAATGGCTCTGCGACGGCTTT 49408
162 GlnSerProCysIleIleGlyGlySerGlyLysGlyLysSerThrIe 178
|||||
49409 CAGAGCCCTCGCATCATTTGAAGGGGAATCGCCAAAGGCAGATGCCACT 49458
178 uLeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysAlaLeu 195
|||||
49459 GCTGCAGCCGATTCGATGCTGGGGCTCCGAAAGTGCAGAGCTCTGA 49508
195 hrlsPheLysPheValPhePheLeuArgLeuSerArgAlaGlnGly 211
|||||
49509 CCAAGTTCAAAATTCGCTCTTCTTCCTCGCTCAGAGGGCCAGGGGTGA 49558
212 LeuPheGluThrLeuCysAspGlnLeuLeuAspIleProGlyThrIleAr 228
|||||
49559 CTTTGGAAACCTCTGTGATCAATCCGGAATATCTGGCACAATTCAG 49608
228 gLysGlnThrPheMetAlaMetLeuLysLeuArgGlnArgValLeuP 245
|||||
49609 GAAGCAGACATTCATGCGCTGCTGAAAGCTGCGGACAGGGGTCTTT 49658
245 heLeuLeuAspGlyTyrasnGluPheLysProGlnAsnGlyProGluIle 261
|||||

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source

Artiguenave, F., Davoine, C.S., Cruaud, C., Durf, A., Wincker, P.,
Brothier, P., Catolico, L., Barbe, V., Burzender, J.M.,
Brud'Homme, J.F., Brice, A., Fontaine, B., Hellig, R. and
Weissenbach, J.
Spastin, a novel AAA protein, is altered in the most frequent form
of autosomal dominant spastic paraplegia
Nat. Genet. (1999) In press
2 (bases 1 to 138909)
Genoscope
Direct Submission
Submitted (29-FEB-2000) to the EMBL/GenBank/DBJ databases
On Mar 6, 2000 this sequence version replaced gi:6002386.
Location/Qualifiers
1. 138909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone_11b="C17B_978_SKB"
/clone="164M19"
/clone="164M19"
/clone="164M19"

BASE COUNT 39243 a 28424 c 29121 g 42121 t
ORIGIN

alignment_scores:
Quality: 3733.00 Length: 1067
Ratio: 4.854 Gaps: 4
Percent Similarity: 72.071 Percent Identity: 70.947

alignment_block:
US-09-697-089-2 x CNS01DS3/rev ..

Align seg 1/1 to reverse of: CNS01DS3 from: 1 to: 138909

```
1 MetAspHe1le1ysAspAsnSerAr1AlaLeu1leG1naArgMetG1yme 17
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
76388 GTGAAATTCATTAAGGACAAATAGCCGACCCCTATTCAAAAGATGGGAAT 76339
17 tThVal1le1ysG1n1leTh1rAspAsp1leuPheVal1TrpAsnVal1eua 34
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
76338 GACTGTTATTAAGCAAAATACAGATGACCTATTGTTATGGAATGTTCTGA 76289
34 snArg1uG1uVal1asn1le1eCysCy1u1ysVal1G1u1nAsp1a 50
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
76288 ATCGCAAGAGATTAACATCATTTGCTGCGAAGAGTGAGCAGGATGCT 76239
51 AlaArg1y1le1le1HisMet1le1eulys1ysG1ySer1uSerCys1as 67
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
76238 GCTAGAGGATCATTCACATGATTTTGAAAAAGGTTCAAGAGTCCTGTAA 76189
67 n1euPhe1eulysSer1eulysG1u1TrpAsn1Ty1rPro1eulPhe1G1naS1pL 84
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
76188 CCTCTTCTTAATCCCTTAAGGAGTGAACTATCTCTATTTCAGGACT 76139
84 euAsnG1y1n..... 87
  ::::::::::
76138 TGAATGACAAAGTAAGTATCAGTTAGTTTGCATCTGTTTCGCAAGTA 76089
87 ..... 87
  ::::::::::
76088 GAATAGACCAACAAGATCATCTTCCCATGCGAGAGCCCTGCTTT 76039
87 ..... 87
  ::::::::::
76038 TCCTTCTTCATATCACCTGCGACGTGGCCTTGATACCTGCTAAGAAA 75989
87 ..... 87
  ::::::::::
75988 AGCTTAGATGATTTTAGAAAAACTATAGATCAATTGATGCTGAAG 75939
87 ..... 87
  ::::::::::
75938 CCAAAACAATAACCCCAAGCCAGGAATAAATGATTGAGTAATTGCTCA 75889
```

```
87 ..... 87
75888 CTGCAAGATAAAACTGGAGATAAGTCACTAGTTAAGCTTCCAGTGCAG 75839
87 ..... 87
75838 CAGTGGTTCAGAGATTCCCTGGTTGGATGAGATCTACATGAGGAGAGA 75789
87 ..... 87
75788 GGCCACAGCTACTACTACTCATGTCAGGAGCCCTGTTAAGAGAGAGCC 75739
87 ..... 87
75738 TGACTACACAGAACCATGATTCTATGTTTCATTAATAATTTGTTTTATT 75689
87 ..... 87
75688 TTTTAAATAGTAATAAGTCATGACAGATTTTACTCAGAGTGAAGGCC 75639
87 ..... 87
75638 TACTATGATGACAATTATCCCTAACCATCTTCTGCCAAATTCCTTAC 75589
87 ..... 87
75588 TTTCCTTAACCCAAAAGAAAGCTCTCCCTTTTCCAGCAGCTGTATTA 75539
87 ..... 87
75538 GATTGGTTGTGAAGAGAGAAACAGAAAGCAAAAGGGGAGTGGGGTA 75489
87 ..... 87
75488 AAGATCTGGGGAAGGGGAGCAGACAGAACCCCTTTGTGCCAGCTCAG 75439
87 ..... 87
75438 AACTGGAAAGACATTACTTACCAAGGAGAGATTTCAGCCTCAGCC 75389
87 ..... 87
75388 CCCCCTCATCTCTGAACTAGTTTTGAGAGACACAGAAATGGGT 75339
87 ..... 87
75338 CTTCCTTAACATCACCTCTCTCTAATAGGTCTTTTCATCACACATCAG 75289
95 luG1yAsp1euaSpAsp1eua1aG1aAsp1eulysAsp1eulys1rHis1Th 111
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
75288 AAGGAGCTTGAGCATTTGGCTCAGAGATTTAAAGACTTGTAACCATACC 75239
112 ProSerPhe1euaSnPhe1Ty1rPro1eulG1y1aSp1leAsp1le1lePh 128
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
75238 CCAATCTTTCTGAACCTTTATCCCTTGGTGAAGATATATGACATTTT 75189
128 euAsn1eulysSer1Th1rPhe1Th1r1uPro1Val1leu1Trp1Arg1y1Asp1G1nH 145
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
75188 TAACTGAAAGACCTTCACAGAACCTGCTGCTGAGAGAGAACCAAC 75139
145 1sh1sh1sArG1Val1G1u1n1eulTh1r1euaSnG1y1eul1n1a1eul 161
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
75138 ACCATCACCGGCTGAGACAGCTGACCTGATGCGCTCTGCGAGGCTCTT 75089
162 G1nSerProCys1le1leG1uG1u1SerG1y1ysG1y1ysSer1rHis 178
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
75088 CAGAGCCCTGCATCATTTGAAGGGAATCGGCAAGGCAAGTCCACTCT 75039
178 ul1euln1Arg1le1a1eul1eulTrp1y1SerG1y1ysCys1y1sAla1eul 195
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::
75038 GCTGACAGCAATTTGCATGCTGCGGGCTCCGGAAGAGTCAAGGGCTCTGA 74989
195 h1r1ysPhe1ysPhe1Val1PhePhe1euln1rG1eulSerAr1aG1nG1y1y 211
```

|||||
74988 CCAAGTCAAAATGCTTCTTCCTCCGCTCAGAGGCGCCAGGCTGGA 74939
212 LeuphegluThrLeucCysAspGlnLeuLeuAspIleProglyThrIleArg 228
|||||
74938 CTTTGTGAACCCCTCTGTGATCAACTCCTGGATATACCTGGACAAATCAG 74889
228 gLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgValLeup 245
74888 GAAGCAGCATTCATGATGATGCTGCTGAAAGCTGCGCAGAGGCTTCTTT 74839
245 helLeuLeuAspGlyTyrAsnGlnPheLysProGlnAsnGlyProGlnIle 261
|||||
74838 TCCTCTGATGAGCTACAAATGATTCAGCCCAAGCTGCCCCAGAAATTC 74789
262 GluAlaLeuLysGlnAsnHisArgPheLysAsnMetValIleValThr 278
74788 GAAGCCCTGATTAAGAAACACCGCTTCAGAAACATGCTATGCTCAG 74739
278 rThrThrThrGlyCysLeuArgHisIleArgGlnPheGlyAlaLeuThr 295
74738 CACTACCACTGATGCTGAGGACATACGGCAGTTGGTGGCTGACTG 74689
295 IagIuValGlyAspMetThrGluAspSerAlaGlnAlaLeuIleArgGlu 311
74688 CTGAGGTGGGGATATGACAGAAAGACCGCCAGGCTCTCATCCGAGAA 74639
312 ValLeuLysGlnLeuAlaGlnGlyLeuLeuGlnIleGlnLysSe 328
|||||
74638 GTGCTGATCAAGAGACTGCTGAAAGCTTGTGCTCAAAATTCAGAAATC 74589
328 rArgCysLeuArgAsnLeuMetLysThrProLeuPheValValIleThr 345
74588 CAGGTGCTTGAGAAATCATGAAAGACCCCTCTTGTGGTCAATCAGCTT 74539
345 ySAIleGlnMetGlyGlnSerGlnPheHisSerHisThrGlnThr 361
74538 GTGCAATCCAGATGGGTGAAAGTGTCCACTCAGACACAAACAACG 74489
362 LeupheHisThrPheTyrAspLeuLysGlnLysAsnLysHisLysHis 378
74488 CTGTTCCATACCTCTATGATCTGTTGATACGAAACAAACAAACAA 74439
378 sLysGlyValAlaAlaSerAspPheIleArgSerLeuAspHisCysGly 395
74438 TAAAGCTGTGCTGCAAGTCACTTCATTCGGAGCTGAGCCACTGTGAG 74389
395 sPLeuAlaLeuGlnGlyValPheSerHisLysPheAspPheGlnGln 411
74388 ACCTAGCTCTGAGGCTGTCTTCTCCCAAGTTTGATTTGCAACTGAG 74339
412 AspValSerSerValAsnGlnAspValLeuLeuThrThrGlyLeuLeu 428
74338 GATGTGTCACAGCTGAATGAGGATGCTGCTGACAACTGGGCTCTCTG 74289
428 sLysTyrThrAlaGlnArgPheLysProLysTyrLysPhePheHisLys 445
74288 TAAATATACAGCTCAAGCTTCAAGCCAAAGATTAATTTCTTCAAGAT 74239
445 erPheGlnGlyThrAlaGlyArgArgLeuSerSerLeuLeuThrSer 461
74238 CATTCACAGAGTACACAGCAGAGACAGCTCAGAGTTTATGACGCT 74189
462 HisGluProGlnGluValThrLysGlnLysAsnGlyTyrLeuGlnLysMet 478
74188 CATGAGCCAGAGAGAGTACCAAGGGAATGTTACTTGCAAGAAATGCT 74139
478 IserIleSerAspIleThrSerThrTyrSerSerLeuLeuArgTyrThr 495
74138 TTTCATTTTCGACATTCATCATTATAGACGCTCTCGGTACACCT 74089
495 ySGlySerSerValGluAlaThrArgAlaValMetLysHisLeuAla 511
|||||

74088 GTGGGTCAATCTGTGGAGCCACCAAGGCTGTATTGAAAGCACTCGCAGCA 74039
512 ValTyrGlnHisGlyCysLeuLeuGlyLeuSerIleAlaLysArgProLe 528
|||||
74038 GTGTATCAACAGCGCTGCCCTTCTGCGACTTTCATCGCCAGAGGCTCT 73939
528 uTPArgGlnGlnSerLeuGlnSerValLysAsnThrThrGlnGlnGlu 545
73988 CTGGAGACAGCATCTTGGCAAGGTGAAAAACACCACTGAGCAAGAAA 73939
545 IeLeuLysAlaIleAsnLysSerPheValGluCysGlyIleHisLeu 561
73938 TTCTGAAAGCCATTAACATCAATTCCTTTGTAGAGTGTGGATCCATTTA 73889
562 TyrGlnGlnSerThrSerLysSerAlaLeuSerGlnGlnPheGluAla 578
73888 TATCAAGAGAGTACATCCAAATCAAGCCCTGAGCCAGAAATTTGAAGCTT 73839
578 erPheGlnLysSerLeuTyrIleAsnSerGlyAsnIleProAspTyrL 595
73838 CTTTCAAGGTAAACCTTATATATCACTCAGAGAACATCCCGATTACT 73789
595 euPheAspPhePheGlnHisLeuProAsnGlyAlaSerAlaLeuAspPhe 611
73788 TATTTGACTTCTTTGACATTTGCCCAATTTGTCAAGTCCCTGGACTTC 73739
612 IleLysLeuAspPheTyrGlyValAlaMetAlaSerTPGluLysAla 628
73738 ATTAACPTGAGCTTTTATGGGGAGCTATGGCTTCATGGAAGAGCTGC 73689
628 AGluAspThrGlyIleHisMetGluGluAlaProGluThrTyrIleP 645
73688 AGAAACACAGGTGGAATCCACATGGAAGAGCCCAAGAACATCAATTC 73639
645 roSerArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThr 661
73638 CCAGAGGCTGTATCTTTGTTCTTCACTGGAAGCAGAAATTCAGGACT 73589
662 LeuGluValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIle 678
73588 CTGAGGTGCACACTCCGGGATTTGAGCAAGTTGAATAGCAAGATATCAG 73539
678 rTyrLeuGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIle 695
73538 ATATCTGGGAAATATTCAGCTCTGCCCAAGCTTCAGCTGCAAAATTA 73489
695 ySArgCysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCys 711
73488 AGAGATGCTGCTGTGGCTGGAAGCTTCAGTTTGCTTCAAGCACTGT 73439
712 LysAsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGlu 728
73438 AAGAACATTTATTTCTCATGTTGGAAGCCAGTCCCTCACATATAGAGA 73389
728 pGluArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis 745
73388 TGAGAGCAGCATCATCTGTACAAACCTGAAACCTTGATATTCATG 73339
745 sPLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLeuGly 761
73338 ACCTACAGAAATCAAGGCTGCCGGGT..... 73313
762 LeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMetAsn.... 776
73312 AFTGAAATATACAGCTGTTGTGCTTGTCTTCACTTAAAAAATAATACAG 73263
777GluGluAspAlaIleLysLeuAlaGluGly 786
73262 TGTATTAATTTGAAAAAGAGAGAGACATTTATTTCTTCTAAAGGT 73214
seq_name: gb_PR:R2005417
seq_documentation_block:

LOCUS IR2005417 1355 bp mRNA PRI 16-JUL-2000
 DEFINITION Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 2005417.
 ACCESSION AL389934
 VERSION AL389934.1 GI:9367839
 KEYWORDS FLI_CDNA.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1355)
 Aufray, C., Ansoorge, W., Ballabio, A., Estivill, X., Gibson, K.,
 Lehrach, H., Poustka, A. and Lundeberg, J.
 The European IMAGE consortium for integrated molecular analysis of
 human gene transcripts
 unpublished
 2 (bases 1 to 1355)
 Pluvinet, R., Estivill, X., Escarceller, M. and Sumoy, L.
 Direct Submission
 Submitted (15-JUL-2000) Dept. Genetica Molecular, Institut de
 Recerca Oncologica (IRO), Hospital Duran i Reynals, Av. Gran Via
 s/n Km 2,7, L'Hospitalet de Llobregat, 08907 Barcelona, Catalunya,
 SPAIN. Tel.: +34-93-260-7775 Fax: +34-93-260-7776 WWW site:
 http://www.iro.es e-mail enquiries: lsumoy@iro.es
 EURO-IMAGE Consortium Contact: Aufray C
 CNRS UPR 420 - Genetique Molculaire et Biologie du Developement
 IFR 1221 - Rue Guy Moquet 19, Batiment G - Bp 8
 94801 Villejuif Cedex, FRANCE
 Tel: ++33-1-49 58 34 98
 Fax: ++33-1-49 58 35 09
 e-mail: aufray@infobiogen.fr
 This clone is available royalty-free through IMAGE Consortium
 Distributors.
 IMPORTANT: This sequence represents the full insert of this IMAGE
 cDNA clone. No attempt has been made to verify whether this
 corresponds to the full-length of the original mRNA from which it
 was derived.

FEATURES
 source Location/Qualifiers
 1..1355
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="IMAGE cDNA clone 2005417"
 /clone_1ib="NCI_CGAP_Paul"
 /map="2p21-p22"
 <1..1321
 /codon_start=2
 /product="hypotheical protein, weakly similar to
 (AF102871) neuronal apoptosis inhibitory protein 2 [mus
 musculus]"
 /protein_id="CAB97523.1"
 /db_xref="GI:9367840"
 /translation="INSGNIPDYLFDFEHLPCNSALDFIKLDIFYGGAASWEKAAL
 DTGILHMEAEPTIIPSAVSLFENWKEFRLEVTLDLDFSKLNKODIRYLGKLFSSA
 TSLRQIKRCAGVAGSLVLTCKNIYSLMVEASPLTEDRHITSVNLKTLISHD
 LONQRLPGGLDLSLGNLKLKILMDINKMEEDAIKAEGLKLNKKKPLHLTHLD
 IGEGRDYIVKSLSEPCDLERETOLVSCLSANAVYIIAONLHNHLYKLSILLSEVYLE
 KDGNALHEHLDKRMVLEFOLTAIAMPNGCDVGSISLKLLEEVPOIVKGIKRMRL
 TDTEIRILGAFERKPKLPKNFOOLNLAGNRVSSDGLAIVGTFENIKQIVLFFDFSTKER
 LPDPALVKKLSQVSKLFTLDGARLVGWDFDDDLVITGAFKLVTAA"

CDS
 BASE COUNT 416 a 265 c 312 g 362 t
 ORIGIN
 alignment_scores:
 Quality: 2244.00 Length: 439
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 99.772 Percent Identity: 99.772
 alignment_block:
 US-09-697-089-2 x IR2005417 ..
 Align seg 1/1 to: IR2005417 from: 1 to: 1355

586 ILeasSercIyAsnIleProasPtyrleuPheaspPhehgluhIste 602
 |||||
 2 ATCAACTCAGGGAACATCCCGATTACTTATTGACTCTTTGAACATTT 51
 |||||
 602 uProasCysAlaSerAlaLeuAspPheIleLysLeuAspPheTyrGlyG 619
 |||||
 52 GCCCAATTGTCGAAGTGCCTGCGACTTCATTAACTGAGACTTTATGGGG 101
 |||||
 619 lYalameAlaSerTPpGluYAlaAlaGluAspThrGlyIleHis 635
 |||||
 102 GACCTATGGCTTCATGGAAAAGCGTCGAGAACACACAGCTGGAATCCAC 151
 |||||
 636 MetGluAlaAlaProGluPheThrYrIleProSerArgAlaValSerleuPh 652
 |||||
 152 ATGGAAGAGGCCCGAGAACCTACATTCCCGAGGCGCTGATCTTGTGT 201
 |||||
 652 ePheasTrpLysGlnGluPheArgThrleuGluValThrleuArgAsp 669
 |||||
 202 CTTCACCTGGAGCAGCAATTTCAGACTCTGGAGTCCACACTCCGGGATT 251
 |||||
 669 heserLysleuAsnLysGlnAspIleThrTyrleuGlyLysIlePheSer 685
 |||||
 252 TCAGCAAGTTGAATTAAGCAAGATATCATCTGGGAAAATATTCACG 301
 |||||
 686 SerAlaThrSerleuArgleuGlnIleLysArgCysAlaGlyValAlaGlu 702
 |||||
 302 TCTGCCACCAAGCTCAGGCTGCATAATAAGACATGTGCTGTGCTGG 351
 |||||
 702 ySerleuSerleuValleuSerThrCysLysAsnIleTyrSerleuMetV 719
 |||||
 352 AACCTCAGTTGTGCTCAGACACTGTAAAGAACATTTATTCCTCATG 401
 |||||
 719 alGluAlaSerProleuThrIleGluAspGluArgHisIleThrSerVal 735
 |||||
 402 TGGAAACCAAGTCCCTCACCATTAAGAGATGAGGCACATCATCTGTA 451
 |||||
 736 ThrAsnleuLysThrleuSerIleHisAspLeuGlnAsnGluArgleuP 752
 |||||
 452 ACNAACCTGAAAACCTTGAGTATTCTGACCTACAAATCAACAGCGTGGC 501
 |||||
 752 ogLylGlyleuThrAspSerleuGlyAsnleuLysAsnleuThrLysleuI 769
 |||||
 502 GGGTGTCTGACTGACAGCTGGGTAACTGGAAGAACCTTACAAACCTCA 551
 |||||
 769 lMetAspAsnIleLysMetAsnGluAspAlaIleLysleuAlaGlu 785
 |||||
 552 TAATGATTAACATTAAGATGAATGAAGAAGATGCTATTAACACTAGCTGAA 601
 |||||
 786 GlLeuLysAsnleuLysLysMetCysleuPheHisLeuThrHisleuSe 802
 |||||
 602 GGCCTGAAAACCTGAAGAAGATGTGTTTATTTCATTGACCCACTGTGC 651
 |||||
 802 rasPIleGlyGluGlyMetAspTyrIleValLysSerleuSerSerGluP 819
 |||||
 652 TGACATTGGAGAGGAATGATGATACATAGCAAGCTCTGTGTAAGTGAA 701
 |||||
 819 rCysAspLeuGluGluIleGlnleuValSerCysLysleuSerAlaAsn 835
 |||||
 702 CCGTATACCTTGAAGAAATTCATTAGTCTCCGCTGCTGCTGCAAAAT 751
 |||||
 836 AlaValLysIleleuAlaGlnAsnleuHisAsnleuValLysleuSerI 852
 |||||
 752 GCAGTGAATAATCTAGCTCAGAACTTTCACATTTTGTGTAACCTAGCAT 801
 |||||
 802 TCTTGAATTATCAGAAATTTACCTGGAAGATGAATGAAGCTCTTC 851
 |||||
 869 isGluIleuIleAspArgMetAsnValleuGluGlnleuThrAlaIleuMet 885
 |||||
 852 ATGAACCTGATCGACAGATGAACGTGCTAGAACAGCTCACCCGACATGAG 901
 |||||
 886 leuProTrpGlyCysAspValGlnGlySerleuSerSerleuLysHis 902

```

|||||
902 CTGCGCTGGGCTGTGACGTGCAGGCGCCGTGACACCGCTGTTGAACA 951
902 sLeuGlulValProGlnLeuValLysLeuGlyLeuLysAsnTParGly 919
902 TTTTGAGAGAGGTCACCAACTGTCACAGCTGGGTTGAAAACTGGAGAC 1001
919 eUrThrAspThrGluIleArg11IleuGlyAlaPhePheGlyLysAsnPro 935
1002 TCACAGATACAGAGATAGAAATTTAGTGCAATTTTGGAAAGAACCT 1051
936 LeuLysAsnPheGlnGlnLeuAsnLeuAlaGlyAsnArgValSerSerAs 952
1052 CTGAAAACTTCACAGCAGTGAATTTGGCGGGAATCGGTAGACAGCA 1101
952 pGlyTrpLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLeuValP 969
1102 TGGATGGCTTGCTTCATGGGTGATTTGAGAACTCTTAAGCATTTAGTGT 1151
969 hepAspSerThrLysGluPheLeuProAspProAlaLeuValArg 985
1152 TTTTGGACTTGTAGTACTAAGAAATTTCTACGATCCAGCATTTAGTCAGA 1201
986 LysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAlaArgLe 1002
1202 AAACCTTACCAAGTGTATCCAAAGTTACTTTCTGCAAGAGAGCTAGGCT 1251
1002 uValGlyTrpGlnPheAspAspAspLeuSerValIleThrGlyAlaP 1019
1252 TGTGGGGGCAATTTGATGATGATGATCTCATGTTTATACAGTGTCTT 1301
1019 heLysLeuValThrAla 1024
1302 TTAACCTAGTACTGCT 1318

seq_name: gb_pr:AY027788

seq_documentation_block:
LOCUS AY027788 1395 bp mRNA PRI 20-JUL-2001
DEFINITION Homo sapiens CLANB (CLAN1) mRNA, complete cds.
ACCESSION AY027788
VERSION AY027788.1 GI:14324114
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Damiano,J.S., Stehlik,C., Plo,F., Godzik,A. and Reed,J.C.
TITLE Genomics. 75 (1-3), 77-83 (2001)
JOURNAL MEDLINE
PUBMED 21365712
11472070
2 (bases 1 to 1395)
Stehlik,C., Damiano,J.S., Plo,F., Godzik,A. and Reed,J.C.
Direct Submission
Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source
1.1395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="lung"
1.1395
/gene="CLAN1"
/277..1356
/gene="CLAN1"
/codon_start=1
/product="CLANB"
/protein_id="AAK14777.1"

```

```

/db_xref="GI:14324115"
/translation="MNFIRDNSKALLIORMGMVYIKOITDLEFWMVNLNREYNIICCE
/KVEODARGLIIMHILKKSGESCNLFKSLKENMYPLFODLNGSGSLDILGMLKNTK
LIMDNIKMNEDEYIKLAEGLKNTKKMKLHLPHLSPIGGMVYIVKSLSEPCDLEEI
OLVSCCLSANVYIILNONLHNLYKLSIILSEVLEKDCNMLAHEIIDRMNVLEDTA
LMLPGCDYOGSLSSILKHLHEEYPOLYVKGKMKMRITDEIRILGAFPEKKNLPKNOO
LNLWGNRVSSDGLWLAEMGVFENKOLVFEFDFSTKEPLPDPAVLRKLSQVLSKLTPLQE
ARLVGMQFFDDDLSTVTYGAFLVTA"
BASE COUNT 436 a 248 c 327 g 384 t
ORIGIN

alignment_scores:
Quality: 1500.50 Length: 1024
Ratio: 4.180 Gaps: 1
Percent Similarity: 35.059 Percent Identity: 35.059

alignment_block:
US-09-697-089-2 x AY027788 ..
Align seg 1/1 to: AY027788 from: 1 to: 1395

1 MetAsnPheIleLysAspAsnSerArgAlaLeuIleGlnArgMetGlyLe 17
277 ATGAATTCATTAAGGACAAATAGCCGACCCCTTATTCAAAGAGGGGAAAT 326
17 TThrValIleLysGlnIleThrAspAspLeuPheValTrpAsnValLeuA 34
327 GACGTGTATTAAGCAATCACAGATGACCTATTGTATGAAATGTTCTGCA 376
34 sNArgGluGluValAsnIleIleCysGlyLysValGluGlnAspAla 50
377 ATCCGGAAGAGTAACATCATTTGCTGCGAAGAGGTGACAGCATGCT 426
51 AlaArgGlyIleIleHisMetIleLeuLysLysGlySerGlySerCysAs 67
427 GCTAAGGGATCATTCACATGATTTTGAAAAAGGTTTCAGAGTCTGTAA 476
67 nLeuPheLeuLysSerLeuLysGluTrpAsnTrpProLeuPheGlnAspL 84
477 CCTGTTCTTAATCCCTTAAGAGATGAGACTATTCCTATTTCAGAGACT 526
84 euAsnGlyGlnSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
527 TGAATGACAAAGT..... 540
101 LeuAlaGlnAspLeuLysAspLeuTrpHisThrProSerPheLeuAsnPh 117
540 ..... 540
117 eTyTrpLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerThrP 134
540 ..... 540
134 heThrGluProValLeuTrpArgLysAspGlnHisHisArgValGlu 150
540 ..... 540
151 GlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuGlnSerProCysIleI 167
540 ..... 540
167 eGluGlyGluSerGlyLysGlyLysSerThrLeuLeuGlnArgIleAla 184
540 ..... 540
184 eLeuTrpGlySerGlyLysCysLysAlaLeuThrLysPheLysPheVal 200
540 ..... 540
201 PhePheLeuArgLeuSerArgAlaGlnGlyGlyLeuPheGluThrLeucy 217
540 ..... 540

```


217 saspcInleuEusplleProGlyThrIleArgLysGlnThrPheMet 234
540 540
234 IamleuLeuLysLeuArgGlnArgValLeuPheLeuLeuAspClyTyr 250
540 540
251 AsnGluPheLysProGlnAsnCysProGluIleGluAlaLeuIleLysGlu 267
540 540
267 uasnHisArgPheLysAsnMetValIleValThrThrThrGluCysL 284
540 540
284 euArgHisIleArgGlnPheGluAlaLeuThrAlaGluValGlyAspMet 300
540 540
301 ThrGluAspSerAlaGlnAlaLeuIleArgGluValLeuIleLysGluLe 317
540 540
317 uAlaGluClyLeuLeuLeuGlnIleGlnLysSerArgCysLeuArgAsnL 334
540 540
334 euMetLysThrProLeuPheValValIleThrCysAlaIleGlnMetCly 350
540 540
351 GluSerGluPheHisSerHisThrGlnThrThrLeuPheHisThrPheTyr 367
540 540
367 rAspLeuLeuIleGlnLysAsnLysHisLysHisLysGlyValAlaAlaAs 384
540 540
384 eAspPheIleArgSerLeuAspHisCysGlyAspLeuAlaLeuGluCly 400
540 540
401 ValPheSerHisLysPheAspPheGluLeuGlnAspValSerSerValAs 417
540 540
417 nGluAspValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGln 434
540 540
434 rGpPheLysProLysTyrLysPhePheHisLysSerPheGlnGluTyrThr 450
540 540
451 AlaGlyArgArgLeuSerSerLeuLeuThrSerHisGluProGluGluVal 467
540 540
467 lThrLysGlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleT 484
540 540
484 hrSerThrTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGlu 500
540 540
501 AlaThrArgAlaValMetLysHisLeuAlaValTyrGlnHisGlyCys 517
540 540

517 sLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnGluSerL 534
540 540
534 euGlnSerValLysAsnThrThrGlnGlnGluIleLeuLysAlaIleAsn 550
540 540
551 IleAsnSerPheValGluCysGlyIleHisLeuTyrGlnGlnSerThrSe 567
540 540
567 rLysSerAlaLeuSerGlnGluPheGluAlaPhePheGlnClyLysSerL 584
540 540
584 euTyrIleAsnSerGlyAsnIleProAspTyrLeuPheAspPheGlu 600
540 540
601 HisLeuProAsnCysAlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
540 540
617 rGlyGlyAlaMetAlaSerTrpGluLysAlaAlaGluAspThrClyGlyI 634
540 540
634 leHisMetGluGluAlaProGluThrTyrIleProSerArgAlaValSer 650
540 540
651 LeuPhePheAsnTrpLysGlnGluPheArgThrLeuGluValThrLeuAr 667
540 540
667 gAspPheSerLysLeuAsnLysGlnAspIleThrTyrLeuGlyLysIleP 684
540 540
684 heSerSerAlaThrSerLeuArgLeuGlnIleLysArgCysAlaGlyVal 700
540 540
701 AlaGlySerLeuSerLeuValLeuSerThrCysLysAsnIleTyrSerLe 717
540 540
717 uMetValGluAlaSerProLeuThrIleGluAspGluArgHisIleThrS 734
540 540
734 eValThrAsnLeuLysThrLeuSerIleHisAspLeuGlnAsnGlnArg 750
540 540
751 LeuProGlyGlyLeuThrAspSerLeuGlyAsnLeuLysAsnLeuThrLy 767
541 541
767 sLeuIleMetAspAsnIleLysMetAsnGluGluAspAlaIleLysLeu 784
582 GCCTATATGATGATACATTAAGATGAATGAAGAAGATGCTATTAAGACTAG 631
784 IacGluGlyLeuLysAsnLeuLysLysMetCysLeuPheHisLeuThrHis 800
632 CTGAAGCGCTGAAGAAACCTGAAGAAGATGCTTATTTCAATTGACCCAC 661
801 LeuSerAspIleGlyGluGluMetAspTyrIleValLysSerLeuSerSe 817
682 TTGCTGTACATTTGGAGGGAATGATTAATAGTCAAGTCTGTCTGTCAG 731
817 rGluProCysAspLeuGluGluIleGlnLeuValSerCysCysLeuSerA 834

```

|||||
732 TCACCCCTGACCTTGAGAAATTCATTAATCTCTCTCTCTCTG 781
834 TAASALValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysIleu 850
782 CAAATGACAGTAAATCCCTAGCTCAGAAATCTTCACATTTGGTCAAACTG 831
851 SerIleLeuAspLeuSerGluAsnTyrLeuGluLysAspGlyAsnGluAl 867
|||||
832 ACCATTCCTGATTATTCAGAAAATTAACCTGGAAAAAGATGGAATGAAAC 881
867 AleuHisGluLeuIleAspArgMetAsnValLeuGluGlnLeuThrAlaL 884
882 TCCTCATGAACTGATCGACAGATGAACTGCTAGAACAGCTCACCGCAC 931
884 eumelleuprotfrpglycysasppvalglnGlySerLeuSerSerLeu 900
932 TCATGCTGCGCTGGGGCTGTGACGTCCAAGCAGCCTGAGACGCTGTG 981
901 LysHisLeuGluGluValProGlnLeuValLysLeuGlyLeuLysAsnTr 917
982 AAACATTTGGAGAGAGTCCACACACTCGTCAAGCTTGGTTGAAAACTG 1031
917 PArgLeuThrAspThrGluIleArgIleLeuGluAlaPhePheGlyLysA 934
1032 GAGACTCAGACAGATACAGAGATTAGATTAGTGCTATTTTGGAAAGA 1081
934 snProLeuLysAsnPhenGlnGlnLeuAsnLeuAlaGlyAsnArgValSer 950
1082 ACCCTCTGAAAACTTCACAGCACTTGATTTGGCGGGAATGCTGTGAC 1131
951 SerAspGlyTyrLeuAlaPheMetGlyValPheGluAsnLeuLysGlnLe 967
1132 ACTGATGATGATGCTTCCCTTCAATGGCTGTATTGAGATCTTAAGCAAT 1181
967 ValAlaPhePheAspPheSerThrLysGluPheLeuProAspProAlaLeu 984
1182 AGTGTTTTGACTTTAGTAAAGAAATTTCTACCTGATCCAGCACTGAG 1231
984 AlArgLysLeuSerGlnValLeuSerLysLeuThrPheLeuGlnGluAla 1000
1232 TCAGAAAACCTTAGCCAGTGTATTCCAAGTTAATCTTCTGCAGAAAGCT 1281
1001 ArgLeuValGlyTyrGlnPheAspAspAspLeuSerValIleThrGln 1017
1282 AGCCTTGTGGGTGGCAATTTGATGATGATTCACAGTGTATTATTCAGG 1331
1017 ValAlaPheLysLeuValThrAla 1024
1332 TGCCTTTAAACTGATTAACCTGCT 1353
-g_name: gb_pr:AY027789
In documentation block:
FEATURES             768 bp      mRNA      PRI      20-JUL-2001
DEFINITION Homo sapiens CLANC (CLANI) mRNA, complete cds.
ACCESSION AY027789
VERSION AY027789.1 GI:14324116
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 768)
AUTHORS Damiano,J.S., Stehlik,C., Plo,F., Godzik,A. and Reed,J.C.
TITLE Clanc, a novel human ced-4-like gene
JOURNAL Genomics. 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
REFERENCE 2 (bases 1 to 768)
AUTHORS Stehlik,C., Damiano,J.S., Plo,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death

```

```

Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
SOURCE              1..768
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="2"
                    /map="2p22-p21"
                    /tissue-type="lung"
gene               1..768
                    /gene="CLANI"
                    277..747
                    /gene="CLANI"
                    /codon_start=1
                    /product="CLANC"
                    /protein_id="AAK14778.1"
                    /db_xref="GI:14324117"
                    /translation="MNFKDNRRALLQRMGMVYIKQITDDLFWYVNLREEVNIICE
                    KVEDDAARGIHMLKKSSCNLFSLKEMVPLFODLNGSLFPHOTSGDLDDIA
                    ODLDLHTEPSFLNFPYLGEDIDIFMLKSTFEPVLMRRDQHHHREQLTLV"
BASE COUNT         218 a      157 c      180 g      213 t
ORIGIN
alignment_scores:
    Quality:      819.00      Length:      154
    Ratio:        5.318      Gaps:        0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-697-089-2 x AY027789
Align seg 1/1 to: AY027789 from: 1 to: 768
1 MetAsnPhelIeLysAspAsnSerArgAlaLeuIleGlnArgMetGlyLe 17
|||||
277 ATGAATTTTATTAAGGACATAGCCGAGCCCTTATTCAAGATGGGAAT 326
17 ThrValIleLysGlnIleThrAspAspLeuPheValTyrPasnValLeu 34
|||||
327 GACTGTTATTAAGCAATCAGATGACTTATTGTATGAAATGTTCTGA 376
34 snArgGluGluValAsnIleIleCysCysGluLysValGlnAspAla 50
|||||
377 ATCGCGAAGAGTAACATCATTTGCTCGCAGAGAGTGAGCAGATGCT 426
51 AlaArgGlyIleIleHisMetIleLeuLysLysGlySerGlySerCysAs 67
|||||
427 GCTAGAGGATCATTCACATGATTTTGAAGAAAGGTTTCAGAGTCTGTAA 476
67 nLeuPheLeuLysSerLeuLysGluTyrPasnTyrTyrPheLeuPheGlnAspL 84
|||||
477 CCTCTTCTTAATCCCTTAAGGAGTGAAGTAATCCTCATTTTCAGAGAT 526
84 euAsnGlyInSerLeuPheHisGlnThrSerGluGlyAspLeuAspAsp 100
|||||
527 TGAATGACAAAGCTTTTTCATCAGACATCAGAAAGAGACTTGGACGAT 576
101 LeuAlaGlnAspLeuLysAspLeuTyrHisThrProSerPheLeuAsnPh 117
|||||
577 TTGGCTCAGAGATTTAAAGGACTTGTACATACCCATCTTTCTGTGAACCT 626
117 eTyrProLeuGlyGluAspIleAspIleIlePheAsnLeuLysSerThr 134
|||||
627 TTATCCCTTGGTGAAGATATGACATTTATTTTAACCTGAAGAGACCT 676
134 heThrGluProValLeuTyrPargLysAspGlnHisHisIсарValGlu 150
|||||
677 TCACAGAACCTGCTCGTGGAGGAAGACCAACACATCATCCGCGGTGAG 726
151 GlnLeuThrLeu 154
|||||
727 CAGCTGACCCCTA 738

```

seq_name: gb_pat:A64531

seq_documentation_block: 6228 bp DNA PAT 29-MAR-1999

LOCUS A64531 Sequence 23 from Patent WO9726331.

DEFINITION A64531

ACCESSION A64531

VERSION A64531.1 GI:3717929

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 6228)

AUTHORS Korneiluk,R.G., Mackenzie,A.E., Roy,N., Robertson,G. and Tama,I.K.

TITLE USE OF NEURONAL APOPTOSIS INHIBITOR PROTEIN (NALP)

JOURNAL Patent: WO 9726331-A 23 24-JUL-1997;

COMMENT UNIV OTTAWA (CA)

FEATURES

Other publication AU 1614997 19970811.

source 1..6228

Location/Qualifiers

BASE COUNT 1825 a 1332 c 1307 g 1764 t

ORIGIN

alignment_scores:

Quality: 655.00 Length: 1108

Ratio: 1.077 Gaps: 42

Percent Similarity: 54.874 Percent Identity: 23.736

alignment_block:

US-09-697-089-2 x A64531 ..

Align seq 1/1 to: A64531 from: 1 to: 6228

69 PheLeuysSerLeuysGluTrpAsnTyProLeupheGlnAspLeuAs 85

1425 TTTTCCAAATATGAAATGTC...TTCGCGAAGTGAATCCAGACCTTCA 1471

85 nglYglN.....SerLeupheHsGlnThrSerGluGlyAspL 98

1472 GAGCGGTGGTGAAGTCTTGTAATTAAGTGAACCAAGTGAAGCAATC 1521

98 euAspAsp..... 100

1522 TTGAAGATTCATATGACAGTGGCTCTTAAGTCCAGAAATGCGACAGGT 1571

101LeuAlaGlnAspLeuLysAs 107

1572 GAAGCCGAGTGTTCAGAGGCAAGAAATCTGAATGACGACGTGACAGC 1621

107 pLeuTyHsIThrProSerPheLeuAsnPhe..... 117

1622 AGCTTAATACAGCGCCAGTTCCGCGACATGTCTTTCCTGATATCTCTT 1671

118TyrProLeuGlyGluAspIleAspIleIlePhe 128

1672 CCGATTCGGCCAGCGACACTGCTGGGCTGTGATCTGTAT..... 1715

129 AsnLeuysSerThrPheThrGluProValLeuThrPargLysAspGlnI 145

1716 ...GCTTCAAACACATCAGCAACCTGTG..... 1742

145 SHSHSHArGValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuG 162

1743CAGAACCTCTGGTGTCTGCTGAGAGTCTTTGGCAACTTGA 1782

162 InSerProGlyIleIleGluGlnGlySerGlyLysSerThrLeu 178

1783 ACTGTGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1832

179 LeuGlnArgIleAlaMetLeuThrPglYserGlyLysAlaLeuTh 195

1833 CTGAGAGAAATAGCTTTCTGTGGGCAATCGATCGTCCCTGTTAA 1882

195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210

1883 CAGGTTCCAGCTGTGTCTTACCTCTCCCTTGTCCACAGCAACAGC 1932

210 LysGlyLeuPheGluThrLeuGlyAspGlnLeuLeuAspIleProGlyThr 226

1933 AGGGGCTGGCCAGTATCATCTGTGACAGCTCTCAGAGAAAGAAAGATCT 1982

227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysLeuArgGlnArgVal 243

1983 GTTACGTGAATGTGCATGAGAAACATTATCCAGAGTTAAAGATCAGGT 2032

243 lLeuPheLeuLeuAspGlyTyrosGluPhe.....LysProGlnAsnc 258

2033 CTATATCTCTTTAGATGATCAAGAAATATGTTCAATCCCTCA.... 2078

258 ysProGluIleGluAlaLeuIleLysGluAsnHsArgPheLysAsnMet 274

2079GTCAATAGGAACCTGATTCAAATAAACCTTATCCCGGACCTGC 2123

275 ValIleValThrThrThrThrGluLysLeuArgHsIleArgGlnPheG 291

2124 CTATGATTCGCTCTCCGTACAAACAGGCGCAGGACATCCGCGATACCT 2173

291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308

2174 AGAGACCATTTCTAGATCCAGACATTCCTTTTATATATCTGTCTGA 2223

308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322

2224 TATTACGGAAGCTCTTTTCAATATATATGACTCTCTGCGAAAGTTATG 2273

323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProle 339

2274 GTTACTTGTGGAAGAACCAAGTTTGCAGAGATACAGAAACCTCTCT 2323

339 upheValValIleThrCysAlaIleGlnMetGlyGluSerGluPheHs 356

2324 CTTTGTGGCGGCGATCTGTCTCATTTGCTTCACTATCTTTGACCCAT 2373

356 eHsIThrGlnThrThrLeuPheHsIThrPheTyAspLeuLeuIleGln 372

2374 CTTTGTGATGATGCTGCTTTTCAAGTCTTATATGAGACGCTTCTCTTA 2423

373 LysAsnLysHsLysHsLysGlyValAlaAlaSerAspPheIleArg.. 388

2424 AGGACAA.....GCGACACTGAAATTCCAAGC 2455

389 SerLeuAspHsCysGlyAspLeuAlaLeuGluGlyValPheSerHsI 405

2456 AACGTGTCTCTCTGTGTGAGTGTGAGCTTGAAGGCTTTTTCATAGTT 2505

405 ysPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419

2506 GCTTTGAGTTTATGATGATGATCTCCAGACAGGAGGTTGATGAAGAT 2555

420 ValLeuLeuThrThrGlyLeuLeuCysLysTyThrAlaGlnArgPhe 436

2556 GAAGATCTAACCATGTCTGATGAGCAAAATTTCACGCCAGAGACTAAG 2605

436 sProLysTyLysPhePheHsLysSerPheGlnGluTyThrAlaGly 453

2606 ACCATTCCTACCGGTTTAAAGTCTCTCCCAAGAAATTTCTTGGGGGA 2655

453 rArgLysSerSerLeuLeuThrSerHsGluProGluGluValThrLys 469

2656 TGAGCTGATTAAGATCTCTGATTCAGATAGCGAGCAACATCAAGATTG 2705

470 GlyAsnGlyTyThrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486

[illegible]


```

4009 .TTGATCTTCCTACTGCGATGGAATTATCGAGTGGCCAACTGATCA 4057
900 euylshlsleugluValProglInleuVallysleuglyleuLysAsn 916
4058 TCCAGCGAGTCTCAGACGCTCATTTGTCGAGTCTCTCATTTTTCACG 4107
917 TrpArgLeuThrAspThr.....GlutIleArgIleLeuG1 928
4108 ACCTTGATCATGACACGCTGCGGAATTGCCAAAGTACATCACTGCG 4157
928 yAlaPhePheGlyLysAsnProLeuLysAsnPheGlnGlnLeuAsnLeuA 945
4158 AGGTTTC.....CAGAACTTGAGAACCTTAAGCTTT 4189
945 IaGlyAsn...ArgValSerSerAspGlyTrpLeuAlaPheMetGlyVal 960
4190 CAATCATCTCAAGATTACAGAGGAGGATACGAAATTTCTTTCAGCA 4239
961 PheGlnAsnLeuLysGlnLeuValPhePheAspPheSer.....Th 974
4240 CTGGACACACTGCGCAATCTTGACGAGTGTGACATCTCCAGCATTTTCA 4289
974 rLysGlnPheLeuProAspProAlaLeuValArgLysLeuSerGlnVal 991
4290 AGAGTGATCAAGCTCAGGCCACACAGTCAGTCTTGAGTCATGTCG 4339
991 euSerLysLeuThrPheLeuGlnGlnAlaArgLeuValGlyTrpGlnPhe 1007
4340 TGTAGACATCACCAAGGCTCATTAAGCTGACATGTTAACTTGCTCTTG 4389
1008 AspAspAspAspLeuSerValIle 1015
4390 GATCGACATGATATGCTGCTTCTT 4413
seq_name: gb_pat:E24989

```

```

seq_documentation_block:
LOCUS   E24989      5984 bp      DNA      PAT      07-SEP-2001
DEFINITION Apoptosis-inhibiting protein, gene encoding the same and cDNA
hereof.
ACCESSION E24989
VERSION   E24989.1 GI:13024687
KEYWORDS  JP 1999116599-A/1.
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 5984)
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS   Shigehiro,I.K.Y.Y.
TITLE     Apoptosis-inhibiting protein, gene encoding the same and cDNA
JOURNAL   Patent: JP 1999116599-A 1 27-APR-1999;
COMMENT
OS       Homo sapiens (human)
PN       JP 1999116599-A/1
PD       27-APR-1999
PF       14-OCT-1997 JP 1997280831
PR
PT       SHIGEHIRO IKEDA,KENJI YAMAMOTO
PC       C07K14/52,C07K16/28,C12N15/09,C12Q1/68//C12P21/02,(C12P21/02,
PC       C12N15/19),
PC       C12N15/00
CC       Strandedness: Double;
CC       Topology: Linear;
FH       key
FT       source
           1. 5984
           /organism='Homo sapiens (human)'.

```

```

FEATURES
source
1. 5984
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 1777 a 1267 c 1267 g 1673 t
ORIGIN

```

```

alignment_scores:
Quality: 653.00      Length: 1108
Ratio: 1.074        Gaps: 42
Percent Similarity: 54.874      Percent Identity: 23.736

```

```

alignment_block:
US-09-697-089-2 x E24989

```

```

Align seg 1/1 to: E24989 from: 1 to: 5984

```

```

69 PheLeuLysSerLeuLysGlnTrpAsnTrpProLeuPheGlnAsnLeuAs 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1321 TTTCTCCAAATATGAAGTCC...TCTGGGGAAGTACTCTCAGACCTTCA 1367
85 nglyGln.....SerLeuPheHlsGlnThrSerGlnGlyAspL 98
1368 GAGCCGCTGTGAACCTTGTGATTAATTAAGTGAACACCAAGTGAAGCAATC 1417
98 euAspAsp..... 100
|||||:|||||
1418 TTGAGATTCAATAGACAGTTGCTATAGTCCAGAAATGACACAGGCT 1467
101 .....LeuAlaGlnAspLeuLysAs 107
1468 GAAGCCGAGTGTTCACAGAGCCAAAGATCTGAATGACGACCTGAGAGC 1517
107 PleuTyrlHisThrProSerPheLeuAsnPhe..... 117
1518 AGCTTATACAGCGCAGTTCCGCCACATGCTTGTGATATGCTCTT 1567
118 .....TyrProLeuGlyLysAspIlePhe 128
1568 CCGATCTGCGCCAGCAGCAGCTCTGCGCTGATCTGTCTAT..... 1611
129 AsnLeuLysSerThrPheThrGlnProValLeuTrpArgLysAspGlnH 145
1612 ...GTTTCAAAACACATCACAAACCTGTG..... 1638
145 shLSHsArgValGluGlnLeuThrLeuAsnGlyLeuLeuGlnAlaLeuG 162
1639 .....CAAGAACCTCTGCTGCTGCTGAGCTGCTTGGCACTTGA 1678
162 InserProCysIleIleGluGlnGluSerGlyLysGlyLysSerTrpLeu 178
1679 ACCTGTGCATGTGTGTGAGAGGTGAAGCTGAAGTGAAGAGAGGTCCTC 1728
179 LeuGlnArgIleAlaMetLeuTrpGlySerGlyLysCysLysAlaLeuTh 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1729 CTGAGAGAAATAGCTTTTCTGTGGCATGTGATGCTGTCCCTGTAA 1778
195 rLysPheLysPheValPhePheLeuArgLeuSer.....ArgAlaGln 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1779 CAGGTTCCAGCTGCTGTTTCTACCTCTCCCTTATGTCACACAGCAGACG 1828
210 LysGlyLeuPheGluTrpLeuLysAspGlnLeuLysIleProGlyThr 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1829 AGGCGCTGCGCAGATCATCTGTGACCACTCTCCTAGAGAAAGAGATCT 1878
227 lLArgLysGlnThrPheMetAlaMetLeuLysLeuLysArgGlnArgVa 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1879 GTTACAGAAATGTGCATGAGCAATATATACACAGCTTAAGATCAGCT 1928
243 lLeuPheLeuLysAspGlyTrpAsnGlnPhe.....LysProGlnAsn 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1929 CTTATTCCTTTTGTAGATGACTCAAGAAATATGTCAATCCCTCA 1974
258 ySProGlnIleGlnAlaLeuIleLysGlnAsnHlsArgPheLysAsnMet 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1975 ....GTCATAGCAAACTGATTCAAATAACACACTTATCCCGAGCTGC 2019
275 ValIleValThrThrThrGluCysLeuArgHlsIleArgGlnPheG1 291

```

```

2020 CTATTGATGTCGTCGTAACAACAGGGCCAGGACATCCGCCGATACCT 2069
2021 YAlaLeuThrAlaGluValAlaGlyAspMetThrGluAspSerAlaGlnAla 308
2070 AGAGACCATTCAGAGATCAAGACATTTCCCTTTATATATACTGCTGTGA 2119
308 euLeaArgGluValLeuLeuGluLeuAla.....GluLeuLeuLeu 322
2120 TATTACGGAGACCTCTTTACATATATATAGCTCCGTCGGAAGTTTATG 2169
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProle 339
2170 GTTATCTTGGAAGACCAAGATTGTGCAAGATACAGAAAACCTCCT 2219
339 uPheValAlaIleThrCysAlaIleGlnMetGlyLysSerGluPheHis 356
2220 CTTTGTGGCGGGGATCTGTGTCATGTTTCAGTATCCTTTGACCCAT 2269
356 eRhIstHcIleThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
2270 CCTTGTGATGATGGCTGTTTCAAGTCCATATATGGAAGCGCTTTCCTTA 2319
373 LysAsnLysHisLysHisLysGlyValAlaAlaAspAspPheIleArg.. 388
2320 AGGACCAAA.....GCGACAGCTGAATTCGAAAGC 2351
389 .SerLeuAspHisCysGlyAspLeuAlaLeuGlnGlyValPheSerHisL 405
2352 AACGTGTGCTCCTGCTGGTGAAGCTGGCTTGAAAGGCTTTTTCATGTT 2401
405 yspHAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
2402 GCTTGTGATTTAATGATGATCATCTCGCAAGACAGGGTTGATGAAGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
2452 GAAGATCTTAACCATGTCTGTGAGCAAAATTACAGCCAGAGACTAAG 2501
436 sProLysTyrLysPhePheHisLysSerPheGlnGlyTyrThrAlaGly 453
2502 ACCATTCTACCGGTTTAAAGTCCTGCCCTCCAGAAATTTCTGGGGGA 2551
453 rGArgLeuSerSerLeuLeuThrSerHisGluProGlnGluValThrLys 469
2552 TGAGGCTATTAAGACTCGTATTCAGATAGCGAGAACATCAAGATTGG 2601
470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
2602 GGACTGTATCATTTGAACAATCACTCAACCATGATGACTGTAAAGCG 2651
486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
2652 CTACAACAATTTTGAACATAATGC.....TCCAGCTCCCTTCAACAA 2695
503 rGAla.....ValMetLysHisLeuAlaValTyrGlnHisGly 516
2696 AAGCAGGGCCCAAAATGTGTCTCATTTGCTCCATTAGTGATGAAC... 2742
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTyrArgGlnLys 533
2743 .....AAAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
2751 ATTGGAGATATATCTGAATAATGATGACTACTTAAGACAGCAGCAAAA 2800
540 ..ThrThrGluGlnGluLeuLeuLysAlaIle.....AsnIle 551
2801 TTTCACTCGCATGCGATTACTTAGGGATGTGTGCAAAATTGTGCACAA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2851 GCTTACTTTTCAATGGTTTCAGAACATTTACTGTCTTCCCTGAAAC 2900
562 ...TyrGlnGlnSerThrSerLysSerAlaLeuSerGlnGluPheGlu 577
2901 TGCTTATTCAAAGCACT...GTTCGTCGGTGTCTTCATTTGTTTGC 2947
577 lAphePheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsp 593
2948 AATTCCTTCAAGGAGAACACAGTACTTTGGGTGCTTAACCTTA...CAG 2994
594 TyrLeuPheAsp..... 597
2995 TACTTTTGACCCACCAGAAAGCTTGTCAATTGTGAGAGATCCACTT 3044
598 .....PhePheGlnHisLeuP 603
3045 CCCATAGAGGAATAAGACATCCAGAGACATTTTTCAGTTCG 3094
603 rOAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
3095 AAACATGTTTGGACAAATCAACAGGTGCCAATATGATCAGAGACTATGCT 3144
618 GlyGly.....AlaMetAlaSerThrGlu..... 625
3145 TCTGCTTTGAACCTATGATGATGAGGAGCGAAATTAGCTGAAAAAGA 3194
626 .....LysAlaAlaGluAspT 631
3195 GGATATATGTAAGACCTATATGATATGACAGCGACGATCACCAGAGC 3244
631 hrGlyGlyIleHisMetGluGluAlaProGluThrTyr...IleProSer 646
3245 TTAGTACTGCTATGTAAGAACTTCTCCAAAGACATCAAGATTCCTCC... 3291
647 ArgAlaValSerLeuPhePheAsnThrLysGlnGluPheArgThrLeuG 663
3292 .....TGTCTAGA 3299
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrL 680
3300 AGTCATGTGATGATATGATTTGATGTGTAGGCCAGAGATGCTTGAGATTC 3349
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
3350 TAATGACAGTTTCTCACCTTCACAGCGCATCGAATCTCAATTTAAACAC 3399
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
3400 AGCAGAGGCTTTATGAAGAAGCATCCGCCAGCTCTGTAGCTGTAAAGC 3449
713 .AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAsp 729
3450 CTCTGTACCAAGTCTCATAGCAAGTTGGAAGTCAAGCCAGCGAGAC 3499
729 lUAArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
3500 AGGAAGCTCTTCTCACCTCCCTCCCTCCGGAATCTTGAAGTCCAGCG 3549
745 .....AspLeuGlnAsnGlnArgLeuProGlnGlyLeuThrAspSerLe 759
3550 ACAATCCAGTCAACAGACCAATCTTCTCTAT.....CTGGATTAAGTT 3593
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet. 775
3594 CCTGTGCTGAAGAAGACTGTCTGTGATCTGTGAGGCGCAATTAATGTTT 3643
775 ..... 775
3644 TTTCAGTATTCCTGGAAGATTTCCAAACTTCCACATATGAGAGAAATTA 3693
776 .....AsnGluGluAspAlaIleLysLeuAlaGluGlyLe 787
3694 TTGATCCAAATTTCAAGCTGATGATGATCTTCCAAACTGATTAATTAAT 3743

```



```

787 ulysasnleuylslysmetCysleupheHisleu.....Thrhisleus 802
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3744 TCMAAAATTCACCAACCTTCATGTTTCCATCTGAACTGAATCTCTTT 3793
      802 eraSpIleGIuGIuGlyMetAspTyrIleValIysSerleuSerSerGlu 818
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3794 CGGATTTTGGGTCTCATGACTATGCTGTTCC..... 3828
      819 Procys...AspleuGIuGIuIleGIuLeuValSerCysCysleuSerAl 834
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3829 ...TGTAAAGAACTCACAGAAATTAAGTTTCGATTCATTTT... 3870
      834 aasnAlaValIysIleleuAlaGlnAsnleuHIsnleuValIysleu 851
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3871 ..CAAGCCGTCCTCCATTGTTGCC...AGTTGCCAAATTTATTCTCTGA 3916
      851 erIleleuAspleu...SerGluAsnTyrleuGIuIysAspGIuAsnGlu 866
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3917 AGATATTAAATCTTGAAGCCAGCAATTCCTGATGAGAAACATCAGAA 3966
      867 AlaIeuHIsleuIleAspArgMetAsnValleuGIuIleuThrAl 883
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3967 AAATTTGCCACATTTTAGCTTCTTAGTAACCTGCAGAA..... 4008
      883 aleuMetleuProTropIyCysAspValGIuIysSerleuSerSerleu 900
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4009 ..TTGATCTCTCTACTGCGGATGGAATTTTCAGTGCGCCAAACGTATCA 4057
      900 eulysHIsleuGIuGIuValProGIuLeuValIysleuGIuLeuLysAsn 916
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4058 TCCAGCAGTGTCAAGACCTTCATGTCGCGAGCTCTCATTTTTCAG 4107
      917 TrpArgleuThrAspThr.....GluIleArgIleleuGI 928
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4108 ACCTTGAATGATGACAGCGGTGGAATTCACAAAGTACAAATCACTGC 4157
      928 yAlaPhePheGIuLysAsnProleuLysAsnPheGIuIleuAsnleu 945
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4158 AGGTTTC.....CAGAAGCTTGAGAACCTTAAGCTTT 4189
      945 laGIuAsn...ArgValSerSerAspIyTrpLeuAlaPheMetGIuVal 960
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4190 CAATTCATATCCAAAGATTACAGAGAGATACAGAAATTTCTTCAAGCA 4239
      961 PheGIuAsnleuLysGIuLeuValPhePheAspPheSer.....Th 974
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4240 CTGGACACACATGGCAACTTGCAGAGATTGCACATCTCCAGCATTTTAC 4289
      974 rIysGIuPheleuProAspProAlaLeuValArgLysleuSerGIuVal 991
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4290 AGAGGTATTAAGCTCAGGCCACACAGCTCAAGTCTTTGATCATGTG 4339
      991 euserIysleuThrPheleuGIuIleuAlaArgleuValGIyTrpGIuPhe 1007
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4340 TGTACACATACCAAGGCTCATTAAGCTAAGACATGTTAAGTGGCTGTG 4389
      1008 AspAspAspAspIleuSerValIle 1015
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4390 GATGCAGATGATATGCTATGCTT 4413
      seq_name: gb_pat:A64529

```

```

AUTHORS Korneluk,R.G., Mackenzie,A.E., Roy,N., Robertson,G. and Tamai,K.
TITLE USE OF NEURONAL APOPTOSIS INHIBITOR PROTEIN (NAIP)
JOURNAL Patent: WO 9726331-A 21 24-JUL-1997;
UNIV OTTAWA (CA)
COMMENT Other Publication AU 1614997 19970811.
FEATURES Location/Qualifiers
source 1..6124
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1803 a 1298 c 1285 g 1738 t
ORIGIN

alignment_scores:
      Quality: 653.00 Length: 1108
      Ratio: 1.074 Gaps: 42
      Percent Similarity: 54.874 Percent Identity: 23.736

alignment_block:
us-09-697-089-2 x A64529 ..

Align seg 1/1 to: A64529 from: 1 to: 6124

69 PheleuLysSerleuysGIuTrpAsnTyrProleuPheGlnAspleuAs 85
   |||||:.....:|||||:
1321 TTTCTCCAAATATGAAGTCC...TCTGCGAGAGTACTCCACACCTTCA 1367
      85 ngIyGln.....SerleuPheHIsGlnThrSerGIuGIuAspL 98
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1368 GACCCTGGGTGAACCTTTGTGAATTACTGCAACCAAGCAAGCAAGCAATC 1417
      98 eulAspAsp..... 100
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1418 TTGAAGATTCAATAGACAGTTGCTCTATAGTCCAGCAATGSCACAGGCT 1467
      101 .....LeuAlaGlnAspleuLysAs 107
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1468 GAAGCCACGTGTTTCAAGAGCAAGAAATCTGAATGAGACAGCTGAGAGC 1517
      107 pleuTyrHIsThrProSerPheleuAsnpe..... 117
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1518 AGCTTATACAGCGCGCATTTCCGCCACATGCTTCTTCAATATCTCTT 1567
      118 .....TyrProleuGIuGIuAspIleleuPhe 128
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1568 CCGATCTGCCACGACGACCACTTCTGCGCTGAGTGTGTATTT..... 1611
      129 AsnleuLysSerThrPheThrGIuProValLeuTrpArgLysAspGlnH 145
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1612 ..GCTTCAAAACATACACCAACCTGTG..... 1638
      145 SHIsHIsArgValGIuGIuLeuThrleuAsnGIuLeuGlnAlaIleuG 162
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1639 .....CAAGAACCTCTGCTGCTGCTGAGCTGAGCTTGGCAACATTTG 1678
      162 InSerProcysIleIleGIuGIuSerGIuLysGIuLysSerThrIleu 178
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1679 ACTGTGTATGTGTGTGTGAGGTGAAGCTGAAGTGAAGAGAGGTCCTC 1728
      179 LeuGlnArgIleAlaMetleuTrpGIuSerGIuLysCysLysAlaIleuTh 195
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1729 CTGAAGAAATATAGCTTTTCTGTGGCATGTGATGTCTCCCTGTATAA 1778
      195 rIysPheLysPheValPhePheleuArgleuSer.....ArgAlaGlnG 210
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1779 CAGGTTCCAGCTGTGTTTCTACCTCTCCCTTAGTGTACACAGACACAGAG 1828
      210 LysGIuLeuPheGIuThrleuCysAspGIuLeuLeuAspIleProGIuTyr 226
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1829 AGGGGCTGGCCAGTATCATCTGTGACCAAGCTCTTAAGAAAGAGAGATCT 1878
      227 IleArgLysGlnThrPheMetAlaMetleuLeuLysleuArgGlnArgVa 243
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

1879 GTTACTGAATGTCATGAGGAACATTATCCAGACGTTAAAGATCAGCT 1928
 243 LLeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsnG 258
 1929 CTATATCCTTTTAACTAGACTACAAAGAAATATGTTCATCCCTCA..... 1974
 258 ySPROGluIleGluAlaLeuIleLysGluAsnHisArgPheLysAsnMet 274
 1975GTTCATGGAACACTGATTCMAAAAAAACACTTATCCCGGACCTGC 2019
 275 ValIleValIleThrThrThrGluCysLeuArgHisIleArgGlnPheG 291
 2020 CTATGATGCTGCTGCTGACAAACAGGCGCCAGGACATCCGCGATACCT 2069
 291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAla 308
 2070 AGAGACCATCTAGACATCAAGCATTTCCCTTATATATAGTCGTCTGA 2119
 308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
 2120 TATTACGGAACTCTTTTCACATATATATAGCTGCTGCGAAAGTTTATG 2169
 323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrPro 339
 2170 GTTACTTGGGAAGAACCAAGTTGCGAGAGATACAGAAACCTCTCT 2219
 339 uPheValValIleThrCysAlaIleGlnMetGlyGluSerGluPheHis 356
 2220 CTTTGTGGCGGCGATCTGCTCATGTTGTTTCAGTATCCTTTTTCAGCC 2269
 356 eRHSThGlnThrThrLeuPheHisThrPheTyrAspLeuLeuIleGln 372
 2270 CCTTGTGATGATGCTGCTTTTCAAGTCTTATGGAAGCGCTTTCCTTA 2319
 373 LysAsnLysHisLysHisLysGlyValAlaIleAspSerPheIleArg.. 388
 2320 AGCAACAA.....GCGACAGCTGAATTTCTAAAGC 2351
 389 .SerLeuAspHisCysGlyAspLeuAlaLeuGlnGlyValPheSerHisL 405
 2352 AACGTGCTCCCTGCTGCTGAGCTGCGCTTAAAGGCTTTTTCATGTT 2401
 405 ySPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
 2402 GCTTGGATTTAATGATGATGATGATCTCGCAGAAAGGCGGTGATGAAGT 2451
 420 ValLeuLeuThrThrGlyLeuLeuCysLysTyrThrAlaGlnArgPhe 436
 2452 GAAGATCTAACCATGCTGTGATGAGCAAAATTACAGCCAGAGACTAAG 2501
 436 sProLysTyrLysPhePheHisLysSerPheGlnGluTyrThrAlaGly 453
 2502 AACATTTCTACGCGTTTAAAGTCCCTCCCAAGAAATTTCTTGGCGGGA 2551
 453 rGArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
 2552 TGAGGCGTATGAACTCCTGGATTCAGATPAGGACAGCAACATCAAGATTG 2601
 470 GlyAsnGlyTyrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
 2602 GGACGTGATCATTTAAACAATCACTACCCAGATGACGTGTAAGCGC 2651
 486 rTyrSerSerLeuLeuArgTyrThrCysGlySerSerValGluAlaThr 503
 2652 CTACACAAATTTTGAACATAGTC.....TCCAGCTCCCTTCAACAA 2695
 503 rGAla.....ValMetLysHisLeuAlaIleValTyrGlnHisGly 516
 2696 AAGCAGGCGCCAAATTTGTCTCATTTGCTCCATTTAGTGAATAC... 2742
 517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuThrArgGlnGluSe 533
 2743AAAGAGTC 2750

533 rLeuGlnSerValLysAsn..... 539
 2751 ATTGAGAAATATATCTGAATAATGATGACTACTTAAGCAGCAGCAGAA 2800
 540 ..ThrThrGlnGlnGluLeuLysAlaIle.....AsnIle 551
 2801 TTTCACCTCAATGCGATTACTTAGGGGATTTGGCAAAATTTGTCCACA 2850
 552 AsnSerPheValGluCysGlyIleHisLeu..... 561
 2851 GCTTACTTTTAAAGTTTTCAGAACATTTACTGTTCTTCCCGGAAAC 2900
 562 ...TyrGlnGluSerThrSerLysSerAlaLeuSerGlnGluPheGln 577
 2901 TGCTTATCAAGCAACACT...GTGCTGCTGTTCTCCATTTGTTTTC 2947
 577 lAphePheGlnGlyLysSerLeuTyrIleAsnSerGlyAsnIleProAsp 593
 2948 AATTCCTTCAAGGAGAACACTGACTTTGGGCGCTTAACTTA...CAG 2994
 594 TyrLeuPheAsp..... 597
 2995 TACTTTTTCGACACCCAGAAAGCTTGTCAATTTGAGAGACATCCACT 3044
 598PhePheGlnHisLeuP 603
 3045 CCCAATACGAGAAATATAGACATCACCCAGACACATTTTCAGTTCTGG 3094
 603 roAsnGly.....AlaSerAlaLeuAspPheIleLysLeuAspPheTyr 617
 3095 AAACATGTTTTCACAAATCACAGGTCGCACTATATGATCAGGACTATGCT 3144
 618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
 3145 TCTGCTTGAACCTATGATGAATGATGGAGCGCAATTAGCTGAAAGAA 3194
 626LysAlaIleGluAspT 631
 3195 GGATTAATGTAAGAGCTATATGATATGACGCGCAGGCAATCCAGCAGC 3244
 631 hngIlyGlyIleHisMetGluGluAlaProGluThrTyr...IleProSer 646
 3245 TTAGTACTGCGCTATGGAACCTTTCTCCAAAGCAGTACAAAGATGCC... 3291
 647 ArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThrLeuG 663
 3292TGCTTGA 3299
 663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyr 680
 3300 AGTCGATGTGATGATGATGATGTTGTGGCCAGGATGTGCTTGAGATTC 3349
 680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
 3350 TTAATGACAGTTTCTCAGCTTCACAGCGCATGAGACTCATTTAAACAC 3399
 697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
 3400 AGCAGAGCGCTTATAGAAAGCATCGCCGACGCTCTGTGACGTCTAAGC 3449
 713 .AsnIleTyrSerLeuMetValGluAlaSerProLeuThrIleGluAsp 729
 3450 CTCTGTACCAAGTCTCCATATAGCAATGTGAACCTCAGCCAGCGCAGC 3499
 729 lUArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
 3500 AGGACCTGCTTCACCTGCTCCCTGGAATCTCTTGAAGTCTTCACAGG 3549
 745AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
 3550 ACAATTCAGTCAACAAGCAAAATCTTCTTAAT.....CTGATTAAGTT 3593

```

* 759 ucglaasnlleuysasnlleuthrlsleuilemetasasnlleuysmet. 775
      : |||||::|||::: |||||::|||:::
3594 CCGTGGCTCAAGAACTGCTGTGATCTGAGGCAATATAATGTTT 3643
775 ..... 775
3644 TTTCAGTATTCCTGAAGATTTCCAAACTTCCACCATATGAGAAATTA 3693
776 ..... 776
3694 TTGATCCAAATTTCACTGATGATGATCTTCCAACTAGTAAATTAAT 3743
787 ..... 787
787 ulysasnlleuysmetcysleupheislu.....Thrlsleus 802
3744 TCAAAATTCGCAAACTTCATCTTTCCATCTGAACTGTAAGTCTTTT 3793
802 eraspilleglgluglmetasprrtlevallyserserleuserglu 818
3794 CGGATTTTGGCTCTCATGACTATGCTTGTTC..... 3828
819 Procys...Aspleuglgluglglleuvalsercyscysleuseral 834
3829 ...TGTAAAGAACTCAGAAATTAAGTTTTCGATTCATTTT..... 3870
834 aasnlalavallyleuvalaglnasnlleuhsasnlleuvallysleus 851
3871 .CAAGCGTCCCATTTGTTGCC...AGTTGCCAAATTTTATTTCTCTGA 3916
851 erlleuysaspleu...Serglusanlyrleuglulysaspjlsanglu 866
3917 AGATATTAAATCTTGAAAGCCACAAATTTCTGATGAGCAATACATGAG 3966
867 Alaleuhsleuileaspargmelastvalleuglglleuvalhrl 883
3967 AAATTTGCTACATTTTGGTCTCTTAGAACCTTGAAAGAA..... 4008
883 aleumelleuprotprglycysaspvalglnglyserleuserleul 900
4009 .TTGATCCTTCTACTGGGATGATGATTTATGAGTGCCAAACTGATG 4057
900 eulshisleuglgluvalproglleuvallysleuglyleuysasn 916
4058 TTCAGCAGTCTCAGCAGCTTCATGTCTCCGAGTCTCATTTTTCAG 4107
917 Trparginleuthrasphr.....Gluilearglleuengl 928
4108 ACTTGAATGATGACAGCGTGGAATTCGCCAAAGTAGCAATCAGTGG 4157
928 yalaphepneglylysasnproleuysasphneglglleuvalasnleu 945
4158 AGGTTTC.....CAGAACTTGAGAACCTTAACCTTT 4189
945 laclaysn...Argvalseraseraspjlytrleualapheketglyval 960
4190 CAATCAATCATCAAGATTACAGAGAGAGATACGAATTTCTTTCAAGCA 4239
961 Phegluasleuysglleuvalphepneaspheaser.....Th 974
4240 CTGGACAACTGCCAAACTTGACAGAGTTGGACATCTCCAGGCATTTTC 4289
974 rlysglupheleuproaspproalaleuvalarglysleuserglval 991
4290 AGAGGTATCAAAAGCTCAGGCCACAAAGTCAAGTCTTGAGTCAATG 4339
991 euserlysleuthrphleuenglglualargleuvalgllytrpglnphe 1007
4340 TGTAGACTACCAAGCTCATTTAGACTGACATGACATTAAGTTGGCTCTG 4389
1008 Aspaspasaspleuseralile 1015
4390 GATGCAGATGATATTGCAATTTGCTT 4413
seq_name: gb_pr:HSU19251

```

```

seq documentation block: 6124 bp mRNA PRI 13-MAR-1998
LOCUS HSU19251 Homo sapiens neuronal apoptosis inhibitory protein mRNA, complete cds.
DEFINITION Homo sapiens neuronal apoptosis inhibitory protein mRNA, complete cds.
ACCESSION U19251
VERSION U19251.1 GI:2642132
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 6124)
Roy,N., Mahadevan,M.S., Mclean,M., Shuter,G., Yaraqli,Z.,
Farahani,R., Tamai,K., Ioannou,P., de Jong,P.J., Ikeda,J.,
Korneluk,R.G., and Mackenzie,A.
The gene for neuronal apoptosis inhibitory protein is partially
deleted in individuals with spinal muscular atrophy
Cell 80 (1), 167-178 (1995)
JOURNAL 95112344
MEDLINE
REFERENCE 2 (bases 1 to 6124)
Chen,Q., Baird,S.D., Besner-Johnston,A., Farahani,R.,
Xuan,J.-Y., Kang,X., Lefebvre,C., Ikeda,J.-E., Korneluk,R.G. and
Mackenzie,A.E.
Sequence of a 131-kb region of 5q13.1 containing the spinal
muscular atrophy candidate genes SMN and N1P
Genomics 48 (1), 121-127 (1998)
JOURNAL 98163755
MEDLINE
REFERENCE 3 (bases 1 to 6124)
Baird,S.D.
Direct Submission
Submitted (29-DEC-1994) Stephen D. Baird, Children's Hospital of
Eastern Ontario, Molecular Genetics, 401 Smyth Rd., Ottawa,
Ontario, K1H 8L1, Canada
4 (bases 1 to 6124)
Baird,S.D.
Direct Submission
Submitted (24-NOV-1997) Stephen D. Baird, Children's Hospital of
Eastern Ontario, Molecular Genetics, 401 Smyth Rd., Ottawa,
Ontario, K1H 8L1, Canada
On Nov 24, 1997 this sequence version replaced gi:632544.
COMMENT
FEATURES
source
exon
exon
exon
exon
exon
exon
CDS

```

```

1. .6124
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5q13.1"
/chromosome="5"
/tissue_type="brain"
/dev_stage="fetus"
1. .73
/number=1
74. .162
/number=2
163. .288
/number=3
289. .859
/number=4
292. .4503
/function="inhibitor of apoptosis"
/note="deleted in people with spinal muscular atrophy SMA;
N1P"
/codon_start=1
/evidence=experimental
/product="neuronal apoptosis inhibitory protein"
/protein_id="AAC52045.1"
/db_xref="GI:2642133"
/translational="MATROKASDERISOPDHNILPELSALIGLDAVOLAKELEEBEOK
ERAKMOGYNSOMSEAKRLKTPYTERPSMWIPEWMAAGFYTGKSGIOFCCSL
ILRGAGTLPRLPIEDHKREHPDCGLMKDKNIAKXIRVKNKSRGKMKYOBE
ARLASFRNPPFYVOGISPCVLSKAGFYFTKODTVQCFSGGCLGNMEBGDDPWKEHA
KWPKCEFLRSKSSSEITIOYIOSYKGFVDITGEHFVNSWVQRELPMASAYCNDSTFA
YEBRLDSFDMRPSAVGVAALAKAGLPYTGKIDIVOCFCGCGCLEKMOEGDDPLDD
HTRCFPNCPLQNMKSSAEVTPDLQSRGELCELLLETTSSESNLBDSIAVGPVEMAG

```



```

: 389 .SerLeuAspHisCysGlyAspLeuAlaLeuGluGlyValPheSerHisL 405
: 2352 AACTGTCTCTCCCTGCTGAGCTGGCCCTGAAGAAGGCTTTTTCATGCT 2401
405 ysrPheAspPheGluLeuGlnAspVal.....SerSerValAsnGluAsp 419
2402 GCCTTGGATTAAATGATGATGATCTGCAGAACAGGAGGCTTGAATGAAGAT 2451
420 ValLeuLeuThrThrGlyLeuLeuCysLysThrThrAlaGlnArgPheL 436
2452 GAAGATCTAACCATGCTGCTGATGAGCAAAATTACACCCAGAGACTAAG 2501
436 sProlysThrLysPhePheHisLysSerPheGlnGluThrThrAlaGlyA 453
2502 ACCATCTACCGGTTTAACTGCTGCTCCAAAGATTCTTGCGGGA 2551
453 rGArgLeuSerSerLeuLeuThrSerHisGluProGluGluValThrLys 469
2552 TGAGGCTGATGAACCTCGGATTGAGTGCAGATGAGCAGAACATCAAGATTGG 2601
470 GlyAsnGlyThrLeuGlnLysMetValSerLysAspLysPheSerThr 486
2602 GGACTGATCATTTGAACAACAACTACCCATGATGACTGTAAAGCGC 2651
486 rTySerSerLeuLeuArgThrThrCysGlySerSerValGluAlaThrA 503
2652 CTACAACAATTTTGAACATGTC.....TCCAGCTCCCTTCAACAA 2695
503 rGAla.....ValMetLysHisLeuAlaValAlaGlyGlnHisGly 516
2696 AACGAGGCCCAAAATGTGTCATGTCATTTGCTCATTTAGTGAGTAAC... 2742
517 CysLeuLeuGlyLeuSerLysLeuAlaLysArgProLeuThrArgGlnGly 533
2743 .....AAGAGTC 2750
533 rLeuGlnSerValLysAsn..... 539
2751 ATTGAGAAATATATCTGAAGAATGATGACTACTTAAGCACACCACAGAAA 2800
540 .ThrThrGlnGlnGluLeuLysAlaLe.....AsnIle 551
2801 TTTCACATGCGATGACGATGACTACTAGGGGATTTGCGCAAAATTTGTCACAA 2850
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
2851 GCCTTACTTTCAATGCTTTCAAGAACTTTACTGCTTCTGCCCTGAAC 2900
562 ....TyGlnGlnSerThrSerLysSerAlaLeuSerGlnGluPheGluA 577
2901 TCGTTATCAAGACACT...GTGCTGCGTGTCTCCATTTGTTTTC 2947
577 IahPhePheGlnGlyLysSerLeuThrIleAsnSerGlnLysIleProAsp 593
2948 AATTCTCTCAAGGAGAACACTGACTTTGGTGGCTTAACTTA...CAG 2994
594 TyrLeuPheAsp..... 597
2995 TACTTTTTCAGCACCCAGAAAGCTTGTCACTTTTGAGAGCATCCACTT 3044
598 .....PhePheGlnHisLeuP 603
3045 CCCAATACGAGAAATAGACATCACCCAGACACATTTTTCAGTTCTGG 3094
603 rAsnGlyCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
3095 AAACATGATTGTTGACAAATCACAGTGCACAACTATGATACGAGACTATGCT 3144
618 GlyGly.....AlaMetAlaSerThrGlu..... 625
3145 TCTGCTTTTAACCTATGAATGAATGGAGACGAAATTTAGCTAAAAAGA 3194
626 .....LysAlaIleAlaGluAspT 631
3195 GCATTAATGTAAGACGCTATATGATATGCAGCCAGGAGCATCACAGACC 3244
631 hrcLysGlyIleHisMetGluGluAlaProGluThrTy...IleProSer 646
3245 TTGTGATCGCTATTTGAAAACCTTCTCAAGACGATCAAGATTC... 3291
647 ArgAlaValSerLeuPhePheAsnTrpLysGlnGluPheArgThrLeuG 663
3292 .....TGCTTAA 3299
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTrL 680
3300 AGTCGATGGAATGATGATGATGATGTAGGCCAGATATGCTGAGATTC 3349
680 euGlyLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
3350 TAATGACAGTTTCTCAGCTTCACAGCCGATGCACTCATTTAAACACAC 3399
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys.. 712
3400 AGCAGAGGCTTTATAGAAAGCATCCGCCACTTGAGCTGTATAGCC 3449
713 .AsnIleTySerLeuMetValGluAlaSerProLeuThrIleGluAspG 729
3450 CTCTGTACACCAAGTCTCATAGCAAGTTGGAACCTCAGCGACGCCGAC 3499
729 IuhArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
3500 AGGAACCTGCTTCACACCTGCTCCCTGGAACTCTGTGAAGCTCAGG 3549
745 .....AspLeuGlnAsnGlnArgLeuProGlyGlyLeuThrAspSerLe 759
3550 ACAATCAGTCACAGACCAATCTTCCATAT.....CTGATTAAGTT 3593
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet.. 775
3594 CCTGTCCTGAAAGAACTGCTGTGATCTGGAGGCAATATTAAGCTTT 3643
775 ..... 775
3644 TTTCAGTCATTCCTGAAGAATTTCCAAACTCCACCATAGAGAAATTA 3693
776 .....AsnGluGluAspAlaIleLysLeuAlaGluGlyLe 787
3694 TTGATCCAAATTTACAGTGAATGATCTCCAAACTAGTAAATTAAT 3743
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeuS 802
3744 TCAAAATTTCCAAACCTTCATGTTTCCATCGAAGTAACTTCTTT 3793
802 eTrAspIleGlyGluGlyMetAspTrpIleValLysSerLeuSerSerGlu 818
3794 CGGATTTGGTCTCATGACTATCTGTTTCC..... 3828
819 ProCys...AspLeuGlnGluIleGlnLeuValSerCysCysLeuSerAl 834
3829 ...TGTAAAGAACTCACAGAAATTAAGTTTTCGATTCATTTT... 3870
834 AsnAlaValLysIleLeuAlaGlnAsnIleHisAsnLeuValLysLeuS 851
3871 .CAGGCGTCCCATTTGTTGCC...AGTTGCAAAATTTATTTCTGCTGA 3916
851 eTrIleLeuAspLeu...SerGluAsnTrpLeuGlnLysAspLysGlnG 866
3917 AGATATTAAATCTTGAAGCCAGCAATTTCTGATGAGAAACATCAGAA 3966
867 AlaLeuHisGluLeuIleAspArgMetAspValLeuGlnGluLeuThrAl 883
3967 AATTGTCCTACATTTTACGTTCTGTACTAACTGGAAAGAA..... 4008
883 AlaMetLeuProTrpGlyCysAspValGlnGlySerLeuSerSerLeuL 900

```

```

4009 . TTGATCTTCTCTACTGGGATGGAATTATTCAGCTGGCCAAATGATCA 4057
900 eulysH1sleuGlulValProGlnleuValIlysleuGlyleuLysasn 916
      :::::::::::::::::::: ||| ::::::::::::::::::::
4058 TTCAGCATGTCCAGCAGCTTCATTCTCTCCAGTCTCTCATTTTTCAG 4107
917 TTPATgIeUthrAspThr.....GluIeArgIleleuG1 928
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
4108 ACTTGATGATGATGACGCGTGGAATTCACAAATGACAAATGACATGCG 4157
928 yAlaPhehheGlyLysAsnProLeuLysAsnPhelGlnIleuAsnLeuA 945
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
4158 AGGTTTC.....CAGAACTTTPAGAACCCTAAAGCTTT 4189
945 lAglyAsp...ArgValserSerAspGlyTTPLeuAlaPheMetGlyVal 960
      :: ||| :::::::::::::::::::: ||::: ||::: ||::: ||:::
4190 CAATCAATACACAAGATTACAGAGAGATACAGAAATTTCTTTCAACCA 4239
961 PheGluAsnLeuLysGlnLeuValPhePheAspPheSer.....Th 974
      :::::::::::::::::::: ||::: ||::: ||::: ||::: ||:::
4240 CTGCAACAATGCCAAACTGCAGAGATTGGACATCTCCAGCATTTTCAC 4289
974 rLysIleuPheLeuProAspProAlaLeuValArgLysLeuSerGlnVal 991
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
4290 AGAGGTATCAAAAGCTCAGGCCACACACAGTCAGCTTTGAGTCATGTG 4339
991 euSerLysLeuThrPheLeuGlnGlnAlaArgLeuValGlyTTPGlnPhe 1007
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
4340 TGTTCAGCATTCACAAAGCTCATTTAGACAGTCAATGTTAGTTGGCTTGG 4389
1008 AspaSpAspAspLeuSerValIle 1015
      ||| ||||| ::::::::::::::::::::
4390 GATCGAGATGATATGTCATTCGCTT 4413
seq_name: gb_pat:A64510

seq_documentation_block:
LOCUS A64510 6133 bp DNA PAT 29-MAR-1999
DEFINITION Sequence 2 from Patent WO9726331.
ACCESSION A64510
VERSION A64510.1 GI:3717909
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Korneluk,R.G., Mackenzie,A.E., Roy,N., Robertson,G. and Tamai,K.
TITLE USE OF NEURONAL APOPTOSIS INHIBITOR PROTEIN (NAIP)
JOURNAL Patent: WO 9726331-A 2 24-JUL-1997;
UNIV OTTAWA (CA)
OTHER PUBLICATION AU 1614997 19970811.
DOCUMENT
FEATURES
source 1..6133
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1806 a 1299 c 1288 g 1740 t
ORIGIN
alignment_scores:
Quality: 653.00 Length: 1108
Ratio: 1.074 Gaps: 42
Percent Similarity: 54.874 Percent Identity: 23.736

alignment_block:
US-09-697-089-2 x A64510 ..
Align seg 1/1 to: A64510 from: 1 to: 6133
69 PheulysSerLeuLysGluThrAsnTyrProLeuPheGlnAspLeuAs 85
||||| :::::::::::::::::::: ||| ::::::::::::::::::::
1330 TTCTCTCCAAATATACAGATCC...TCTCGGAGAGTACTCCAGACCTTCA 1376

```

```

85 nGlyIle.....SerLeuPheH1sGlnThrSerGluLysPL 98
      :::::::::::::::::::: ||||| ::::::::::::::::::::
1377 GAGCCGCTGTGAACCTTGTGAATTACTGGAACCAACCAAGTAAAGCAATC 1426
98 euAspAsp..... 100
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1427 TTGAAGATTCAATAGCAGTTGCTCTATATGTCCAGAAATGGCACAGGCT 1476
101 .....LeuAlaGlnAspLeuLysAs 107
      |||
1477 GAAGCCCACTGCTTTCAAGAGGCAAAAGATCTGAATGACAGCTGAGAC 1536
107 PLeuTyrH1sThrProSerPheLeuAsnPhe..... 117
      ||| :: ||||| :::::
1527 AGCTATATCCAGCGCACTTCCGCCACATGCTTGTGTTGATATCTCTT 1576
118 .....TyrProLeuGlyGluAspIleAspIlePhe 128
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1577 CCGATCTGGCCACGACCATCTGCTGGCTGTGATCTCTATT..... 1620
129 AsnLeuLysSerThrPheThrGluProValLeuThrArgLysAspGlnH1 145
      :::::::::::::::::::: |||||
1621 ..GCTTAAACACATACGCAAACTGTG..... 1647
145 sH1sH1sArgValGlnGlnLeuThrLeuAsnGlyLeuGlnAlaLeuG 162
      ||| ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1648 .....CAAGAACCTCTGTGCTGCTCGTAGAGTCTTTGGCAACTTGA 1687
162 InseTProcysIleIleGlnGlyLysSerGlyLysGlyLysSerThrLeu 178
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1688 ACTGTGCATGTGTGTGAGGAGTGAGAGCTGGAAGTGAAGACGCTGCTC 1737
179 LeuGlnArgIleAlaMetLeuThrIleGlySerGlyLysCysLysAlaLeuTh 195
      :::::::::::::::::::: ||||| ||::: ||::: ||::: ||:::
1738 CTGAAGAATAATACCTTTCTCTGCGCATCTGATGCTGCTCCCTGTAAA 1787
1788 CAGGTTCCAGCTGTTTCTACCTCTCCCTTAATGCCACAGCACAGACG 1837
210 LysGlyLeuPheGluThrLeuLysAspGlnLeuAspIleProGlyThr 226
      ||||| ::::: ||||| ||||| ||||| ||||| ||||| ||:::
1838 AGGGCTGTGGCCAGTATCATCTGTGACAGCTCTAGAGAAAGAGAGACT 1887
227 IleArgLysGlnThrPheMetAlaMetLeuLeuLysArgGlnArgVal 243
      :::::::::::::::::::: ||::: ||::: ||::: ||::: ||::: ||:::
1888 GTTACTGAAATGTGCATCAGACATATATCCAGCAGTTAAAGATTCAGGT 1937
243 lLeuPheLeuLeuAspGlyTyrAsnGluPhe.....LysProGlnAsnC 258
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1938 CTATATTCCTTTAGATGACTCAAAAGAAATATGTTCATCCCTCA... 1983
258 ySPROGluIleGluAlaLeuIleLysGluAsnH1sArgPheLysAsnMet 274
      ||| ||||| :::::::::::::::::::: ||::: ||::: ||::: ||:::
1984 .....GTCAATGGAAACTGATTCAAAAAACCACTATATCCCGGACCTGC 2028
275 ValIleValThrThrThrThrGluCysLeuArgH1sIleArgGlnPheG1 291
      :::::::::::::::::::: ||::: ||::: ||||| ||||| ||::: ||:::
2029 CTATTTGATTGCTGTCCGTACAAACAGGCGCCAGGACATCCGCCATACCT 2078
291 yAlaLeuThrAlaGluValGlyAspMetThrGluAspSerAlaGlnAlaL 308
      ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
2079 AGAGACCATTTCTAGAGATCAAAAGCATTTCCCTTTATATATCTGTGTA 2128
308 euIleArgGluValLeuIleLysGluLeuAla.....GluGlyLeuLeu 322
      ::::: ||||| ::::: ||::: ||::: ||::: ||::: ||::: ||:::
2129 TATTACGGAAGCTCTTTTCACATATATAGACTGCTGCGAAAGCTTATG 2178
323 LeuGlnIleGlnLysSerArgCysLeuArgAsnLeuMetLysThrProLe 339
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
2179 GTTACTTGTGAAGAACCAAGATTGTGAGAAAGATACGAAAAACTCTCT 2228
339 uPheValValIleThrCysAlaIleGlnMetGlyLysSerGluPheH1s 356

```

```

2229 CTTTGTGGCGGCGATCTGCTCATTTGCTTTCAGTATCCTTTGACCAT 2278
      |||||||: |||||: |||
356 erhsthrnglnthrleuphehsthrphetyraspleuleuilegln 372
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2279 CCTTGTAGTATGCGCTTTTCAAGCTATATGAAAGCGCTTTCCTTA 2328
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
373 LysAsnLysHisLysHisLysGlyValAlaLysAspPheIlearg 388
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2329 AGCAACAA.....GCGACAGCTGAATTTCTCAAGC 2360
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
389 ..SerLeuAspHisCysGlyAspLeuAlaLeuGlnGluValPheSerHisL 405
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2361 AACTGATCTCTCTGCTGCTGAGCTGGCTTGAAGGCTTTTTCATGTT 2410
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
405 yspheAspPheGlnLeuGlnAspVal.....SerSerValAsnGluAsp 419
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2411 GCCTTGAATTAAATGATGATGATCGCACAGACAGGGGTTGATGAAGAT 2460
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
420 ValLeuLeuThrThrGlyLeuLeuCysLysSerThrAlaGlnArgPheLy 436
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2461 GAAGATCTAACCATGCTGCTGATGAGCAAAATTACACCCAGACACTAAG 2510
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 sProlysThrLysPhePheHisLysSerPheGlnGluThrThrAlaGlyA 453
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2511 ACCATCTACCGGGTTTAACTCTGCTGCCCTTCAAGAAATTTCTGGCGGA 2560
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
453 rgrArgLeuSerSerLeuLeuThrSerHisGluProGlnGluValThrLys 469
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2561 TGAAGCGGATGAACTCTGGATTCAGATAGGACGAAACATCAAGATTGG 2610
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
470 GlyAsnGlyThrLeuGlnLysMetValSerIleSerAspIleThrSerTh 486
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2611 GGACTGATGATTTTGAACAATCACTCCACCATGATGACTGTAAGCCG 2660
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
486 rTySerSerLeuLeuArgTyThrCysGlySerSerValGluAlaThrA 503
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2661 CTACACAAATTTTGTGAACATATGTC.....TCCAGCCTCCCTTCAACAA 2704
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
503 rglAla.....ValMetLysHisLeuAlaAlaValTyrglnHisGly 516
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2705 AACGAGGCCCAAAATTTGTCTCATTTGCTCCATTGATGAGTAAC... 2751
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
517 CysLeuLeuGlyLeuSerIleAlaLysArgProLeuTrpArgGlnGluSe 533
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2752 .....AAAGAGTC 2759
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
533 rLeuGlnSerValLysAsn..... 539
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2760 ATTGGAGAAATATATCTGAATAATGATGACTTAAAGCACCCAGCAAA 2809
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
540 ..ThrThrGlnGlnGluIleLeuLysAlaIle.....AsnIle 551
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2810 TTTCACTGCGAGATGCAAGTACTTAGGGGATTTGGCAAAATTTGTCCAAA 2859
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 AsnSerPheValGluCysGlyIleHisLeu..... 561
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2860 GCCTACTTTTCATGCTTTCAGAACATTTACTGCTTCTTCCCTGGAAGC 2909
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
562 ....TyrglnGluSerThrSerLysSerAlaLeuSerGlnGluPheGlnA 577
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2910 TCGCTTATCAAGCAACACT...CTTGCCTGCTGCTCCATTGTTGTTTGC 2956
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
577 lAhePheGlnGlyLysSerLeuTyrlLeuSerGlnLysAsnIleProAsp 593
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2957 AATTCCCTTCAAGGAGAACACTGACTTTGGTGCGCTTAACCTTA...CAG 3003
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
594 TyrLeuPheAsp..... 597
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3004 TACTTTTTCACACACCCAGAAAGCTTGTCACTTGTGAGAGCATCCACTT 3053
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
598 .....PhePheGlnHisLeup 603
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3054 CCCAATACGAGAAATACACATCACCCAGACACATTTTTCAGTTCTGG 3103
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
603 roAsnCys.....AlaSerAlaLeuAspPheIleLysLeuAspPheTy 617
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3104 AAACATGTTTTCACAAATCACAGGTGCCAATATGATACAGACTATGCT 3153
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
618 GlyGly.....AlaMetAlaSerTrpGlu..... 625
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3154 TCTGCCCTTTCAGACCTATGAAATGAATGGAGCGCAATTTTACGTAAGAA 3203
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
626 .....LysAlaAlaGluAspT 631
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3204 GGATTAATGTAAAGCATATATGATATGACAGCCAGGAGCATCCAGAGCC 3253
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
631 hrGlyGlyIleHisMetGlnGluAlaProGlnThrTyrlleProSer 646
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3254 TTGATCTGCGCTATTTGAAACTTTCTCCAAAGCAGTACAGATTTCC... 3300
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
647 ArgAlaValSerLeuPheAsnTrpLysGlnGluPheArgThrLeuG 663
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3301 .....TGTCCTAGA 3308
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 uValThrLeuArgAspPheSerLysLeuAsnLysGlnAspIleThrTyrl 680
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3309 AGTCGATGTCGATGATGATGATGATGATGATGATGATGATGATGATG 3358
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
680 euGlnLysIlePheSerSerAlaThrSerLeuArgLeuGlnIleLysArg 696
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3359 TAAATGACAGTTTCTCAGCTTCACAGCGCATCGAATCTCATTTTAAAC 3408
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
697 CysAlaGlyValAlaGlySerLeuSerLeuValLeuSerThrCysLys... 712
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3409 AGCAGAGCGCTTTATAGAAAGCATCCGCCAGCTTGATGACTGCTAAGC 3458
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
713 ..AsnIleTySerLeuMetValGluAlaSerProLeuThrIleGluAsp 729
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3459 CTCTGTCACCAAGTCTCCATTAAGCAAGTTGGAACCTCAGCGCAGCCG 3508
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
729 LuArgHisIleThrSerValThrAsnLeuLysThrLeuSerIleHis... 744
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3509 AGGAACTGCTTCTCACCCTGCTCCCTCCGGAATCTCTTGAAGCTCAGG 3558
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
745 .....AspLeuGlnAsnGlnArgLeuProGlnGlyLeuThrAspSer 759
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3559 ACAATCCAGTCACACAGACCAATCTTCCAT...CTGATAGATT 3602
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
759 uGlyAsnLeuLysAsnLeuThrLysLeuIleMetAspAsnIleLysMet 775
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3603 CCTGTGCTGCAAGAACTGTCTGTGATCTGGAGGCAATATTAATGTTT 3652
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
775 ..... 775
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3653 TTTCACTCATTTCTGAGAAATTTCCAAACTTCCACCATATGAGAAATTA 3702
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
776 .....AsnGlnGluAspAlaIleLysLeuAlaGluGlyLe 787
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3703 TTGATCCAAATTTCACTGATGATGATGATGATGATGATGATGATGATG 3752
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
787 uLysAsnLeuLysLysMetCysLeuPheHisLeu.....ThrHisLeu 802
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3753 TCAAAATTTCTCCAAACCTTCAATGTTTTCATCTGAGAGTAACTTCTT 3802
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
802 erAspIleGlyGlnGlyMetAspTyrlValLysSerLeuSerSerGlu 818
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3803 CGGATTTTGGGTCTCATGACTATGCTGTGTTTC..... 3837
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
819 ProCys...AspLeuGlnGluIleGlnValSerCysLeuSerAla 834
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3838 ...TGTAAAGAACTCACAGAAATTAAGTTTTCGATTCATTTT... 3879
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
834 aAsnAlaValLysIleLeuAlaGlnAsnLeuHisAsnLeuValLysLeu 851
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3880 .CAAGCGTCCCATTTGTTGCC...AGTTTGCCCAATTTTATTTCTCTGA 3925
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 10:50:16 ; Search time 97.58 seconds

(without alignments)
7271.327 Million cell updates/sec

Title: US-09-697-089-1

Perfect score: 3133
Sequence: 1 cgcctaccccggtggaag.....aagttactcgaagccagta 3133

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCRTUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	1.9	5502	3	US-08-836-134-1
2	39.6	1.3	2862	4	US-08-569-749-13
3	39.6	1.3	2862	5	PCT-US96-12860-13
4	39.6	1.3	3151	3	US-09-212-971-13
5	39.6	1.3	3151	3	US-08-800-929A-13
6	39.6	1.3	3151	4	US-09-617-053A-13
7	36.8	1.2	2589	5	PCT-US95-05922A-1
8	36.8	1.2	2589	4	US-08-569-749-1
9	36.8	1.2	2589	5	PCT-US96-12860-1
10	36.8	1.2	3532	3	US-09-205-204-1
11	36.8	1.2	3732	3	US-09-212-971-7
12	36.8	1.2	3732	3	US-08-800-929A-7
13	36.8	1.2	3732	4	US-09-617-053A-7
14	35.2	1.1	2580	2	US-08-511-485-7
15	34.8	1.1	1840	2	US-08-484-938-15
16	34.8	1.1	1840	2	US-08-484-158B-15
17	34.8	1.1	1840	2	US-08-484-596A-15
18	34.8	1.1	1840	2	US-08-480-150A-15
19	34.8	1.1	1840	3	US-08-458-731-15
20	34.8	1.1	1840	3	US-08-149-223A-15
21	34.8	1.1	53526	3	US-08-658-136-2
22	34.8	1.1	53577	3	US-08-658-136-1
23	34.2	1.1	510	2	US-08-934-959-7
24	34.2	1.1	7218	1	US-08-232-463-14
25	34	1.1	200	2	US-08-875-972-27
26	34	1.1	1753	6	5225348-2
27	34	1.1	4695	6	5225348-3

28	33.8	1.1	292	1	US-08-644-664B-41	Sequence 41, Appl
29	33.8	1.1	292	2	US-08-761-277A-41	Sequence 41, Appl
30	33.8	1.1	3979	4	US-09-180-439-1	Sequence 1, Appl
31	33.8	1.1	3979	4	US-09-180-439-2	Sequence 2, Appl
32	33.8	1.1	4123	4	US-09-180-439-7	Sequence 7, Appl
33	33.4	1.1	893	4	US-09-142-320-3	Sequence 3, Appl
34	33.4	1.1	893	4	US-09-142-320-8	Sequence 8, Appl
35	33.4	1.1	1359	2	US-08-851-088-1	Sequence 1, Appl
36	33.4	1.1	3573	4	US-09-353-585-4	Sequence 4, Appl
37	33.4	1.1	5355	1	US-08-089-755A-1	Sequence 1, Appl
38	33.4	1.1	5355	1	US-08-089-755A-4	Sequence 4, Appl
39	33.4	1.1	5355	1	US-08-421-754-1	Sequence 1, Appl
40	33.4	1.1	5535	1	US-08-421-754-4	Sequence 4, Appl
41	33.4	1.1	5535	2	US-08-421-791-1	Sequence 1, Appl
42	33.4	1.1	5535	2	US-08-421-791-4	Sequence 4, Appl
43	33.4	1.1	6471	4	US-09-353-585-1	Sequence 1, Appl
44	33.2	1.1	3624	1	US-07-951-715A-6	Sequence 6, Appl
45	33.2	1.1	3624	2	US-08-459-448A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-836-134-1
; Sequence 1, Application US/08836134A
; Patent No. 6020127
; GENERAL INFORMATION:
; APPLICANT: Mackenzie, Alex E.
; APPLICANT: Korneluk, Robert G.
; APPLICANT: Mahadevan, Mani S.
; APPLICANT: Mclean, Michael
; APPLICANT: Roy, Natalie
; APPLICANT: Ikeda, John
; TITLE OF INVENTION: Neuronal Apoptosis Inhibitor Protein, Gene Sequence and Patent No. 6020127
; TITLE OF INVENTION: Mutations Causative of Spinal Muscular Atrophy
; FILE REFERENCE: 3477-112, 033477/139914
; CURRENT APPLICATION NUMBER: US/08/836,134A
; CURRENT FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5502
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-836-134-1

Query Match 1.9%; Score 59; DB 3; Length 5502;
Best Local Similarity 54.3%; Pred. No. 3.5e-07;
Matches 144; Conservative 0; Mismatches 115; Indels 6; Gaps 1;

QY	535	ttgaagggaatcgcgaagcaagtcacactcgtgcacgcatcgtcgtcggg	594	ttgaagggaatcgcgaagcaagtcacactcgtgcacgcatcgtcgtcggg
DB	1798	tggaggggtgaagcgcggaagtggaagacggtccctccgaagaataatgcttttcgtgg	1857	ttgaagggaatcgcgaagcaagtcacactcgtgcacgcatcgtcgtcggg
QY	595	gtctcggaaggtgcaaggtctgacaaagttcaaatcgtctctcttcctcgtcag--	652	gtctcggaaggtgcaaggtctgacaaagttcaaatcgtctctctctcgtcag--
DB	1858	catctgagatgctgctccctgttaaacaggtccagcgtgttctctacctcctccctagtt	1917	gtctcggaaggtgcaaggtctgacaaagttcaaatcgtctctctctcgtcag--
QY	653	---caaggcccgagggcgttttgaaacccctcttgatcaactcgtgatctactg	708	---caaggcccgagggcgttttgaaacccctcttgatcaactcgtgatctactg
DB	1918	ccaccagaccaggaagggcgtgcccagatcactcttgaccagctctcctagaaagaag	1977	ccaccagaccaggaagggcgtgcccagatcactcttgaccagctctcctagaaagaag
QY	709	gcacaatcaggaagcaacattcattgcatgctgcgcgaagctcgcgcgaaggttctt	768	gcacaatcaggaagcaacattcattgcatgctgcgcgaagctcgcgcgaaggttctt
DB	1978	gatctgttactgaatgtgcatgtgaggaacattacacagcagttaaagaatcaggtctat	2037	gatctgttactgaatgtgcatgtgaggaacattacacagcagttaaagaatcaggtctat
QY	769	tcctcttgatgctgaatgaat 793		tcctcttgatgctgaatgaat 793
DB	2038	tcctttagatgactacaagaat 2062		tcctttagatgactacaagaat 2062

APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
FILE REFERENCE: 07891/009002
CURRENT APPLICATION NUMBER: US/09/212,971B
CURRENT FILING DATE: 1998-12-16
EARLIER APPLICATION NUMBER: 60/017,354
EARLIER FILING DATE: 1996-04-26
EARLIER APPLICATION NUMBER: 60/030,590
EARLIER FILING DATE: 1996-11-14
EARLIER APPLICATION NUMBER: 08/800,929
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 13
LENGTH: 3151
TYPE: DNA
ORGANISM: Mus musculus
US-09-212-971-13

Query Match 1.3%; Score 39.6; DB 3; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.17;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 85 tgactgtataaagcaatcacagatgacctatttgatggaatgtctgaatcgcgaag 144
DB 2175 tgacacatgctcctcctatccttgataatcctcttgaggcagtgtaattacaaaacagg 2234
QY 145 aagtaaacatcatctgctgagaaagtgagcagatgctgctgtagagggatcaccaca 204
DB 2235 aacatgatatattagcagaaacacacagatacccttaacgaagcagagctattgaca 2294
QY 205 tgatttgaaaagggtcagagtcctgtaacctcttcttaaacctcctaaggagtga 264
DB 2295 ccgttttagtcaagggaatgctgcagcagcacaatcttcaaaactctctgaaggaaatg 2354
QY 265 acatcctcatttcagagacttgatgacaagaagctcttctcagactcgaagagag 324
DB 2355 acctcaggtatatagaacttatttgtagaadaagaatagaatataatccacagaag 2414
QY 325 acttgagacattggctcaga 346
DB 2415 acgtttcagcgtctgctatgga 2436

RESULT 5

US-08-800-929A-13
Sequence 13, Application US/08800929A
Patent No. 6133437
GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G
APPLICANT: Mackenzie, Alexander E
APPLICANT: Liston, Peter
APPLICANT: Baird, Stephen
APPLICANT: Tsang, Benjamin K
APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Clark & Elding LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,929A
FILING DATE: 13-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,590
FILING DATE: 14-NOV-1996
APPLICATION NUMBER: 60/017,354
FILING DATE: 26-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bleker-Brady, Kristina
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 07891/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3151 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-800-929A-13

Query Match 1.3%; Score 39.6; DB 3; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.17;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 85 tgactgtataaagcaatcacagatgacctatttgatggaatgtctgaatcgcgaag 144
DB 2175 tgacacatgctcctcctatccttgataatcctcttgaggcagtgtaattacaaaacagg 2234
QY 145 aagtaaacatcatctgctgagaaagtgagcagatgctgctgtagagggatcaccaca 204
DB 2235 AACATGATATTATTAGACAGAAACACAGATACCCCTTACAAAGCAGAGAGCTTATTGACA 2294
QY 205 tgatttgaaaagggtcagagtcctgtaacctcttcttaaacctcctaaggagtga 264
DB 2295 CCGTTTATGTCAGGAAATGCTCAGCCCAACATCTTCAAAACTCTCTGAGGAAATTG 2354
QY 265 acatcctcatttcagagacttgatgacaagaagctcttctcagactcgaagagag 324
DB 2355 ACTCCAGCTTATATGAACCTTATTGTGGAAAAGAAATGAAATGATATTCACACAGAG 2414
QY 325 acttgagacattggctcaga 346
DB 2415 ACGTTTCAGGCTTGTCTATTGGA 2436

RESULT 6

US-09-617-053A-13
Sequence 13, Application US/09617053A
Patent No. 6300492
GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G
APPLICANT: Mackenzie, Alexander E
APPLICANT: Liston, Peter
APPLICANT: Baird, Stephen
APPLICANT: Tsang, Benjamin K
APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
FILE REFERENCE: 07891/009003
CURRENT APPLICATION NUMBER: US/09/617,053A
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/800,929
PRIOR FILING DATE: 1997-02-13
NUMBER OF SEQ ID NOS: 17

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13
LENGTH: 3151
TYPE: DNA
ORGANISM: Mus musculus
US-09-617-053A-13

Query Match 1.3%; Score 39.6; DB 4; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.17;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 85 tgacgtatataaagaacatcacagatgaccttttgtatgaaatgtctgaatcgcaag 144
DB 2175 tgacacatgcttcctcctcctcctgataaactcttctgagcgagtgtaattacaacagc 2234
145 aagtaaacatcatctgcgcgagaagtgagcagatgctcctgaaggagcatcaca 204
DB 2235 aacatgatatataatgacagaaacacacagatcaccttaacagaagaagcttattgaca 2294
QY 205 tgatttgaanaaagggttcagagctccgtbaacctcttcttaaccttaagagtgga 264
DB 2293 cagtttagtcaaggagaatgctgcagcacaacatcttcaaaacctctgaaggaattg 2354
QY 265 aataccctcttctcagacttgatgacaaagcttcttcatcagaatcagaagaag 324
DB 2335 actcaacgtataatgaacttatttgaaagaataatgatataltccaacagaag 2414
QY 325 acttgacgatttgctcaga 346
DB 2415 acgttcagctgtgcatgga 2436

RESULT 7

PCT-US95-05922A-1
Sequence 1, Application PC/TUS9505922A
GENERAL INFORMATION:

APPLICANT: HE, ET AL.
TITLE OF INVENTION: Human Inhibitor of Apoptosis Gene 1
NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/05922A

FILING DATE: 11 MAY 1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 325800-292

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1435 BASE PAIRS

TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR

MOLECULE TYPE: CDNA
PCT-US95-05922A-1

Query Match 1.2%; Score 36.8; DB 5; Length 1435;
Best Local Similarity 46.2%; Pred. No. 0.71;
Matches 122; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 83 aatgactgtataaagaacatcacagatgacctatttgtatgaaatgtctgaatcgca 142
DB 882 attgacatgtgctgcttccatctcctgatatatttttaagggccatgtatttaataa 941
QY 143 agaagtaaacatcatcttgcgcgagaagtgagcagatgctcctagaaggatcatca 202
DB 942 ggaacatgatatatttaataacaaacacacagatcttcaacagacagaaactgatgga 1001
QY 203 catgtatttgaanaaagggttcagagctccgtbaacctcttcttaaccttaagagag 262
DB 1002 tacatttttggttaaggaatgctgcgcaacacatcttcaaaactgtcttaaaagaaat 1061
QY 263 gaactacccctcatctcagacttgatgacaaagcttcttcatcagaatcagaag 322
DB 1062 tgaactgtacattgtatnaaacttatttggatnaagatattgactatattccacaga 1121
QY 323 agacttgacgatttgctcaga 346
DB 1122 agatgtttcagctgtctgacatgga 1145

RESULT 8

US-08-569-749-1
Sequence 1, Application US/08569749

PATENT NO. 6187557

GENERAL INFORMATION:

APPLICANT: Rothe, Mike

APPLICANT: Goeddel, David V

TITLE OF INVENTION: INHIBITORS OF APOPTOSIS

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/569,749

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Brezner, David J.

REGISTRATION NUMBER: 24,774

REFERENCE/DOCKET NUMBER: A-62464/DJB

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415)781-1989

TELEFAX: (415)396-3249

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2589 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-569-749-1

Query Match 1.2%; Score 36.8; DB 4; Length 2589;
Best Local Similarity 46.2%; Pred. No. 1;


```

? APPLICANT: Liston, Peter
? APPLICANT: Baird, Stephen
? APPLICANT: Tsang, Benjamin K
? APPLICANT: Pratt, Christine
? TITLE OF INVENTION: DETECTION AND MODULATION OF IAPs AND
? TITLE OF INVENTION: NAIp FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
? TITLE OF INVENTION: DISEASE
? FILE REFERENCE: 07891/009002
? CURRENT APPLICATION NUMBER: US/09/212,971B
? CURRENT FILING DATE: 1998-12-16
? EARLIER APPLICATION NUMBER: 60/017,354
? EARLIER FILING DATE: 1996-04-26
? EARLIER APPLICATION NUMBER: 60/030,590
? EARLIER FILING DATE: 1996-11-14
? EARLIER APPLICATION NUMBER: 08/800,929
? EARLIER FILING DATE: 1997-02-13
? NUMBER OF SEQ ID NOS: 17
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 7
? LENGTH: 3732
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-212-971-7

```


PR 27-OCT-1999; 99US-0161822.

XX (MILL-) MILLENNIUM PHARM INC.

XX Bertin J, Robison KE;

XX WPI: 2001-308628/32.

XX P-PSDB; AAU02880.

XX Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease.

XX Claim 2; Fig 1; 93pp; English.

CC The sequence represents a cDNA which encodes the human caspase
CC recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
CC number of proteins that transmit signals that activate apoptosis and
CC inflammatory pathways in response to stress and other stimuli. Therefore,
CC CARD-12 and its corresponding nucleic acid may be used in treatment and
CC diagnosis of patients suffering from disorders associated with an
CC abnormal level (an increase or a decrease) of apoptotic cell death or
CC abnormal activity of stress-related pathways. The disorders include
CC cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
CC autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
CC neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
CC sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial
CC infarction, stroke), inflammatory and immune system disorders (e.g.
CC Crohn's disease, insulin-dependent diabetes, contact dermatitis,
CC psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
CC lepromatous leprosy), ischaemic and hypoxic brain injury, kidney
CC ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
CC meningitis and liver disease.

XX Sequence 3133 BP: 903 A; 691 C; 729 G; 810 T; 0 other:

Query Match 100.0%; Score 3133; DB 22; Length 3133;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcctagcccgctgggaagcttcacacgaagaatgaattcaataaggaataagc 60
DB 1 cgcctagcccgctgggaagcttcacacgaagaatgaattcaataaggaataagc 60
QY 61 gagccctatcaaaagaaatggaatgactgtatataaagaatcaagatgacctattg 120
DB 61 gagccctatcaaaagaaatggaatgactgtatataaagaatcaagatgacctattg 120
QY 121 tatggaatgtctgtaacgcgaagaataacatcatctgtcgcgaagaagtgcagcag 180
DB 121 tatggaatgtctgtaacgcgaagaataacatcatctgtcgcgaagaagtgcagcag 180
QY 181 atgcgcgaaggagatcttcacatgatttgaaaaagggttcagaagtccttaacct 240
DB 181 atgcgcgaaggagatcttcacatgatttgaaaaagggttcagaagtccttaacct 240
QY 241 tctctaattccctaaagatggaactccctatcttcaaggaacttgatgacaagtc 300
DB 241 tctctaattccctaaagatggaactccctatcttcaaggaacttgatgacaagtc 300
QY 301 ttttcatcagacatcaagaagagacttgagacatttgctcaggaatttaagaagactgt 360
DB 301 ttttcatcagacatcaagaagagacttgagacatttgctcaggaatttaagaagactgt 360
QY 361 accatacccatcttctgaactttatccctctgtggaagatatgaacattatttta 420
DB 361 accatacccatcttctgaactttatccctctgtggaagatatgaacattatttta 420
QY 421 acttgaagaagacacttcacagaactgtcctgtggaagaagaacaacacacacacgcg 480
DB 421 acttgaagaagacacttcacagaactgtcctgtggaagaagaacaacacacacacgcg 480

QY 481 tggagcagctgacccctgaatggcctcctgcaagctcttcaagagccctgcacatctgaag 540
DB 481 tggagcagctgacccctgaatggcctcctgcaagctcttcaagagccctgcacatctgaag 540
QY 541 gggaaatctggcaaaaggcaagtcacactgtcgcagcgcattgcacatgtctggctccg 600
DB 541 gggaaatctggcaaaaggcaagtcacactgtcgcagcgcattgcacatgtctggctccg 600
QY 601 gaaatgcaagccttggaacaaagttcaaatctgtcttctcctcctgcacagagcc 660
DB 601 gaaatgcaagccttggaacaaagttcaaatctgtcttctcctcctgcacagagcc 660
QY 661 aggtggaacttttgaacccctctgtgacacactcctgatatatacctggaacaatcagga 720
DB 661 aggtggaacttttgaacccctctgtgacacactcctgatatatacctggaacaatcagga 720
QY 721 agcagaacattatggccatgctgctggaagctcggcagaagggcttcttctcttgatg 780
DB 721 agcagaacattatggccatgctgctggaagctcggcagaagggcttcttctcttgatg 780
QY 781 gctacaatgaattcaagcccccgaactgcccagaatcgaagccctgataaaggaaac 840
DB 781 gctacaatgaattcaagcccccgaactgcccagaatcgaagccctgataaaggaaac 840
QY 841 accgcttcaagacatggtlcatcgtcacacacacacacacacacacacacacacacacac 900
DB 841 accgcttcaagacatggtlcatcgtcacacacacacacacacacacacacacacacacac 900
QY 901 agttgtgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 960
DB 901 agttgtgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 960
QY 961 tccgaagaatgtctgacgaagagctgctggaagcttctgttcccaattcagaatcca 1020
DB 961 tccgaagaatgtctgacgaagagctgctggaagcttctgttcccaattcagaatcca 1020
QY 1021 ggtcttgaagaaatctcatgaagacccctctcttgtgtgcatcactgtgcaatcaga 1080
DB 1021 ggtcttgaagaaatctcatgaagacccctctcttgtgtgcatcactgtgcaatcaga 1080
QY 1081 tgggtgaaagtgtgacttcaacttcaacacacacacacacacacacacacacacacacac 1140
DB 1081 tgggtgaaagtgtgacttcaacttcaacacacacacacacacacacacacacacacacac 1140
QY 1141 tgttgaacagaac 1200
DB 1141 tgttgaacagaac 1200
QY 1201 gcttggaacacatgtggaagactagctctggaaggtgtgttctccacaagtatttcg 1260
DB 1201 gcttggaacacatgtggaagactagctctggaaggtgtgttctccacaagtatttcg 1260
QY 1261 aacttcagagatgtgtccagcggtgaatggaatgtccctgtgcaacatggccctctta 1320
DB 1261 aacttcagagatgtgtccagcggtgaatggaatgtccctgtgcaacatggccctctta 1320
QY 1321 aatatacagctcaaaaggttcaagccaagtataaattcttcaacagtcattccagaggt 1380
DB 1321 aatatacagctcaaaaggttcaagccaagtataaattcttcaacagtcattccagaggt 1380
QY 1381 acacagcaggaagcagactcagcagtttatgaagtcacatgagcagagagagagca 1440
DB 1381 acacagcaggaagcagactcagcagtttatgaagtcacatgagcagagagagagca 1440
QY 1441 aggggaatggttacttgcagaagaatgttccatttgcagatatacaccactataaga 1500
DB 1441 aggggaatggttacttgcagaagaatgttccatttgcagatatacaccactataaga 1500
QY 1501 gctctgcctgcagacacctgtgtgcatctgtgtgaagcacaacagggctgttatgaagc 1560
DB 1501 gctctgcctgcagacacctgtgtgcatctgtgtgaagcacaacagggctgttatgaagc 1560

QY 1561 tcgcagcagtgatcaacacacgcgtcctctcggacttccatcgccaagagcctctc 1620
|||||
Db 1561 tcgcagcagtgatcaacacacgcgtcctctcggacttccatcgccaagagcctctc 1620
QY 1621 ggagacaggaattcttgcgaagtgtgaaaacacacactgagcaagaaattctgaaagcca 1680
|||||
Db 1621 ggagacaggaattcttgcgaagtgtgaaaacacacactgagcaagaaattctgaaagcca 1680
QY 1681 taacatcatattccttctgtagagtggtgcacatcttatatacagaagatcatcccaat 1740
|||||
Db 1681 taacatcatattccttctgtagagtggtgcacatcttatatacagaagatcatcccaat 1740
QY 1741 cagcccccagacccaagaatttgaagcttctctcaaggttaaaagcttatatacactcag 1800
|||||
Db 1741 cagcccccagacccaagaatttgaagcttctctcaaggttaaaagcttatatacactcag 1800
QY 1801 ggaacatcccgatctacttacttacttacttacttacttacttacttacttacttactt 1860
|||||
Db 1801 ggaacatcccgatctacttacttacttacttacttacttacttacttacttacttactt 1860
QY 1861 tggacttcatataactgacttacttacttacttacttacttacttacttacttacttactt 1920
|||||
Db 1861 tggacttcatataactgacttacttacttacttacttacttacttacttacttacttactt 1920
QY 1921 aagacacaggtggaatccacatggaagagccccaagaacactacatcccgacaggtg 1980
|||||
Db 1921 aagacacaggtggaatccacatggaagagccccaagaacactacatcccgacaggtg 1980
QY 1981 taatttcttctcaacttgaagcagaagaattcaagactctggaagtgcaactccggagt 2040
|||||
Db 1981 taatttcttctcaacttgaagcagaagaattcaagactctggaagtgcaactccggagt 2040
QY 2041 tcagcaaggtggaatgaagaagatacatatcttggggaaataattcaagctctgccaac 2100
|||||
Db 2041 tcagcaaggtggaatgaagaagatacatatcttggggaaataattcaagctctgccaac 2100
QY 2101 gccctcaggtcgaataaagaagatgctgtgtgtgctggaagcctcaggttgtgtcctca 2160
|||||
Db 2101 gccctcaggtcgaataaagaagatgctgtgtgtgctggaagcctcaggttgtgtcctca 2160
QY 2161 gcaacctgtaagaacattatctctcatgtgtggaagccagtcctccacatagaagtgt 2220
|||||
Db 2161 gcaacctgtaagaacattatctctcatgtgtggaagccagtcctccacatagaagtgt 2220
QY 2221 agagggacatcatcatctgttaacaaactgaaacacttgagatcatcatgaactcaagaatc 2280
|||||
Db 2221 agagggacatcatcatctgttaacaaactgaaacacttgagatcatcatgaactcaagaatc 2280
QY 2281 aacggctgcgggtgtgtctgtgacgaagcttgggttaacttgaagaacacttaagaagctca 2340
|||||
Db 2281 aacggctgcgggtgtgtctgtgacgaagcttgggttaacttgaagaacacttaagaagctca 2340
QY 2341 taatggaatacaataaagaatgaagaagaatgctataaaaactagctgaaagcctgtaaaa 2400
|||||
Db 2341 taatggaatacaataaagaatgaagaagaatgctataaaaactagctgaaagcctgtaaaa 2400
QY 2401 accctgaagaagaatgtgttatttatttatttatttatttatttatttatttatttattt 2460
|||||
Db 2401 accctgaagaagaatgtgttatttatttatttatttatttatttatttatttatttattt 2460
QY 2461 attacatagtcgaagctctctgtcgaagtgaaacacttgacacttgaagaatattcaattagctc 2520
|||||
Db 2461 attacatagtcgaagctctctgtcgaagtgaaacacttgacacttgaagaatattcaattagctc 2520
QY 2521 cctgtctgtctgtcgaagaatgaaatccttagctcagaatcttcaaatcttgatca 2580
|||||
Db 2521 cctgtctgtctgtcgaagaatgaaatccttagctcagaatcttcaaatcttgatca 2580
QY 2581 aacgtgagatctcttgattatcaagaataattaccttgaagaagaatgaagaatgagctctc 2640
|||||
Db 2581 aacgtgagatctcttgattatcaagaataattaccttgaagaagaatgaagaatgagctctc 2640
QY 2641 atgagactatcgacaggtgtagcgtgtagaacagctcacccagctgagctgtgctcctgg 2700

Db 2641 atgagactatcgacaggtgtagcgtgtagaacagctcacccagctgagctgtgctcctgg 2700
|||||
QY 2701 gctgtgacgtgcaagcagccttgagcagcctgttgaacatttgaagaagttcccaaac 2760
|||||
Db 2701 gctgtgacgtgcaagcagccttgagcagcctgttgaacatttgaagaagttcccaaac 2760
QY 2761 tcgtcaagcttgggttgaagaacactggaagactcaagatcacagaagatagaaatttaggtg 2820
|||||
Db 2761 tcgtcaagcttgggttgaagaacactggaagactcaagatcacagaagatagaaatttaggtg 2820
QY 2821 catttttgaagaacaccccttgaaaaaactccagcagcttgaatttggcgggaatcgtg 2880
|||||
Db 2821 catttttgaagaacaccccttgaaaaaactccagcagcttgaatttggcgggaatcgtg 2880
QY 2881 tgaagcagtgatgagtggtgtccttcaatggaatggaatggaatggaatggaatggaatgga 2940
|||||
Db 2881 tgaagcagtgatgagtggtgtccttcaatggaatggaatggaatggaatggaatggaatgga 2940
QY 2941 tttttagcttagtactaaagaatttctactgataccagcattgtagaagaatttagcc 3000
|||||
Db 2941 tttttagcttagtactaaagaatttctactgataccagcattgtagaagaatttagcc 3000
QY 3001 aaggttataccaaacttaacttcttctgcaagaagcttaggtgtgtggtggaattgagt 3060
|||||
Db 3001 aaggttataccaaacttaacttcttctgcaagaagcttaggtgtgtggtggaattgagt 3060
QY 3061 atgatgatctcagtgattatcaaggtgtcctttaaactgaactgaactgtcttaataagtga 3120
|||||
Db 3061 atgatgatctcagtgattatcaaggtgtcctttaaactgaactgaactgtcttaataagtga 3120
QY 3121 ctggaagccagta 3133
|||||
Db 3121 ctggaagccagta 3133

RESULT 2
AAH98254
ID AAH98254 standard; cDNA: 3545 BP.
XX
AC AAH98254;
XX
DT 12-OCT-2001 (first entry)
XX
DE Murine EST-derived coding sequence SEQ ID NO: 111.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Mus musculus.
XX
PN W0200154477-A2.
XX
PD 02-AUG-2001.
XX
FE 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HXYSE-) HXYSE INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI: 2001-476164/51.
DR P-PSDB: AAM23595.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising

Dh	2058	ggaactcaataaactvggctcttataivggggagctabgctctcaivggaaaaagctcgaga	2117
Qy	1922	agacacaggtcvgaaatccacaatgaaagagccccagaaacctcaatctccacagacgctgt	1981
Dh	2118	agacacaggtcvgaaatccacaatgaaagagccccagaaacctcaatctccacagacgctgt	2177
Qy	1982	atcttcttcttccaactvgaaagcaggaatctcagacctctgaggtcacacctccggatctt	2041
Dh	2178	atcttcttcttccaactvgaaagcaggaatctcagacctctgaggtcacacctccggatctt	2237
Qy	2042	caggaagtctgaaatgaagaataatacaatctccatctcgggaaaaaatattcaactctcgacaag	2101
Dh	2238	caggaagtctgaaatgaagaataatacaatctccatctcgggaaaaaatattcaactctcgacaag	2297
Qy	2102	ccctacaggtctcaaataaagagatctcgtctgtgagctgaaagccctcaattgtgtccacg	2161
Dh	2298	ccctacaggtctcaaataaagagatctcgtctgtgagctgaaagccctcaattgtgtccacg	2357
Qy	2162	caactgttaagaacaattatctctctcaatggtggaagccagctccctccacataagaatga	2221
Dh	2358	caactgttaagaacaattatctctctcaatggtggaagccagctccctccacataagaatga	2417
Qy	2222	gagggacatacaactctgttaacaaaccttgaaaaacctgtgattctcaatgacctacagaatca	2281
Dh	2418	gagggacatacaactctgttaacaaaccttgaaaaacctgtgattctcaatgacctacagaatca	2477
Qy	2282	acggctctccgggtgtgtctgtgactacagctctgtgttaactggaagaaaccttacaagctcat	2341
Dh	2478	acggctctccgggtgtgtctgtgactacagctctgtgttaactggaagaaaccttacaagctcat	2537
Qy	2342	aatggaataacataaagaatgaaatgaagaagaatgctabaaaaactagcttgaagcctgaaaaa	2401
Dh	2538	aatggaataacataaagaatgaaatgaagaagaatgctabaaaaactagcttgaagcctgaaaaa	2597
Qy	2402	ccgtgaagaagaatgttgtttatttcaatttgaccacactgtgtcaatctgaaagggaaatgga	2461
Dh	2598	ccgtgaagaagaatgttgtttatttcaatttgaccacactgtgtcaatctgaaagggaaatgga	2637
Qy	2462	ttacatgtcaaaatctctctgtgcaagtgaacccgtgagcccttgaaagaattcaatagtctc	2521
Dh	2658	ttacatgtcaaaatctctctgtgcaagtgaacccgtgagcccttgaaagaattcaatagtctc	2717
Qy	2522	ctgtgctgtctgtctgcaaatgcaatgcaaaatctcagctcagaatctcttcaaatgtgttcaa	2581
Dh	2718	ctgtgctgtctgtctgcaaatgcaatgcaaaatctcagctcagaatctcttcaaatgtgttcaa	2777
Qy	2582	actgaagacatctctgattcttaacagaanaattacactggaanaaagctgaaatgaagctcttca	2641
Dh	2778	actgaagacatctctgattcttaacagaanaattacactggaanaaagctgaaatgaagctcttca	2837
Qy	2642	tgaagctatcgaacagagatgaaagctgtctagaacaagctccacacgacgtgagctcgccctggg	2701
Dh	2838	tgaagctatcgaacagagatgaaagctgtctagaacaagctccacacgacgtgagctcgccctggg	2897
Qy	2702	ctgtgacgtgcaaaagcagccctgagcagccctgttgaanaacttvgagaagagttccacaact	2761
Dh	2898	ctgtgacgtgcaaaagcagccctgagcagccctgttgaanaacttvgagaagagttccacaact	2957
Qy	2762	cgtaaaagcttggtgttgaanaaaactgagagactcaacagatacagaatgaaatltttagtgtgc	2821
Dh	2958	cgtaaaagcttggtgttgaanaaaactgagagactcaacagatacagaatgaaatltttagtgtgc	3017
Qy	2832	attcttttggaaaaagaaaccttgaaaaacttccacagacgttgaatttggcgggaaatcgtgt	2881
Dh	3018	attcttttggaaaaagaaaccttggaaaaacttccacagacgttgaatttggcgggaaatcgtgt	3077
Qy	2882	gagcagtgatgaaatggtcttgccttcaatgggtgtatttggaaatcttaagcaattagtggt	2941
Dh	3078	gagcagtgatgaaatggtcttgccttcaatgggtgtatttggaaatcttaagcaattagtggt	3137
Qy	2942	ttttgactttagttaaagaataattctacccgtatccagacatagtctcagaaaaacttagca	3001
Dh	3138	ttttgactttagttaaagaataattctacccgtatccagacatagtctcagaaaaacttagca	3197

QY	3002	agcgtatcccaagtaactcttctcgcaagaagcctaggctgttgagtgcaattgata	3061
Db	3198	agtgttatcccaagtaactcttctcgcaagaagcctaggctgttgagtgcaattgatga	3257
QY	3062	tgatatactcagtggtttatcaaggtgcctttaactcagtaacgcgttaataaggttac	3121
Db	3258	tgatatactcagtggtttatcaaggtgcctttaactcagtaacgcgttaataaggttac	3317
QY	3122	tcgaagccagta	3133
Db	3318	tcgaagccagta	3329
RESULT	3		
AA503946			
ID	AA503946	standard; DNA; 3615 BP.	
XX	AA503946;		
AC			
XX	12-SEP-2001	(first entry)	
DT			
DE	Human caspase recruitment domain 12 (CARD-12) genomic DNA.		
XX			
KM	Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;		
KM	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;		
KM	systemic lupus erythematosus; arthritis; neurological disorder; stroke;		
KM	Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease		
KM	aplastic anaemia; myocardial infarction; inflammatory disorder;		
KM	Crohn's disease; insulin-dependent diabetes; contact dermatitis;		
KM	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;		
KM	tuberculosis; ischemic brain injury; hypoxic brain injury; ds;		
KM	kidney ischaemia; reperfusion injury; acute bacterial meningitis;		
KM	excitotoxic brain damage; liver disease.		
XX			
OS	Homo sapiens.		
FH			
FT	Key	Location/Qualifiers	
FT	CDS	1..3615	
FT		/*tag= a	
FT		/product= "Human CARD-12"	
XX			
PN	WO200130971-A2.		
XX			
PD	03-MAY-2001.		
XX			
PF	26-OCT-2000; 2000WO-US29643.		
XX			
PR	27-OCT-1999; 99US-0161822.		
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
XX			
PI	Bertin J, Robison KE;		
XX			
DR	WPI; 2001-308628/32.		
XX	P-PSDB; AAU02881.		
PT			
PT	Isolated caspase recruitment domain-12 polypeptide and nucleic acids		
PT	encoding them, useful for treating and diagnosing disorders associated		
PT	with abnormal apoptosis such as cancer, arthritis and Alzheimer's		
PT	disease -		
XX			
PS	Disclosure; Fig 2; 93pp; English.		
XX			
XX			
CC	The sequence represents a genomic DNA which encodes the human caspase		
CC	recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a		
CC	number of proteins that transmit signals that activate apoptosis and		
CC	inflammatory pathways in response to stress and other stimuli. Therefore		
CC	CARD-12 and its corresponding nucleic acid may be used in treatment and		
CC	diagnosis of patients suffering from disorders associated with an		
CC	abnormal level (an increase or a decrease) of apoptotic cell death or		
CC	abnormal activity of stress-related pathways. The disorders include		
CC	cancer, viral infections (e.g. caused by poxviruses, adenoviruses),		

CC autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
CC neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
CC sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial
CC infarction, stroke), inflammatory and immune system disorders (e.g.
CC Crohn's disease, insulin-dependent diabetes, contact dermatitis,
CC psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
CC dermatomus leprosy), ischemic and hypoxic brain injury, kidney
CC ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
CC meningitis and liver disease.
XX
XX Sequence 3615 BP: 1041 A; 811 C; 845 G; 918 T; 0 other;

QY	882	ccctgataaaggaataaccacgcgtctcaagaacaatgylcatcgtccactaccactgaat	882
Db	1259	ccctgataaaggaataaccacgcgtctcaagaacaatgylcatcgtccactaccactgaat	1316
QY	883	gcctgaagcacatccgagcttctgctccctgactctgagtgaggggataatgacagaag	942
Db	1319	gcctgaagcacatccgagcttctgctccctgactctgagtgaggggataatgacagaag	1378
QY	943	acaagcgccagagctctacatccgaanaatgctgtgataagaagcttgctgaagctctgtac	1002
Db	1379	acaagcgccagagctctacatccgaanaatgctgtgataagaagcttgctgaagctctgtac	1438

Query Match	95.1%	Score 2980.6;	DB 22;	Length 3615;
Best Local Similarly	98.1%	Pred. No. 0;		
Matches 3047; Conservative	0;	Mismatches	4;	Indels 54; Gaps 1;

Accession	Sequence	Length
U003	tccaaattccgaattccaggctgacttaggaattcctgaagacctctctcttgggtca	108
Db	tccaaattccgaattccaggctgacttaggaattcctgaagacctctctcttgggtca	1498

QY	37	tgaatttcataaagagaataaagcccgagcccttatcttaaaagatbgygaatgactggtataa	96
	419	tgaatttcataaagagaataaagcccgagcccttatcttaaaagatbgygaatgactggtataa	478
	97	agcaaatcacagatgaccttatctgtaatgagatgttccgaatcgcgagaagtaacaatca	156
Db	479	agcaaatcacagatgaccttatctgtaatgagatgttccgaatcgcgagaagtaacaatca	538
QY	157	ttctgcgcgagaagctgagcagagatgctgctagagaggaatcatcatcatgatttggaaa	216
Db	539	tttctgcgcgagaagctgagcagagatgctgctagagaggaatcatcatcatgatttggaaa	598
QY	217	agggtttcaagagctccctgtaacctctctcttcaaacctcttaagagatgagaaatctctcat	276
Db	599	agggtttcaagagctccctgtaacctctctcttcaaacctcttaagagatgagaaatctctcat	658
QY	277	ttcagagacttgaatcagcaaaa-----	297
Db	659	ttcagagacttgaatcagcaaaagtcttgaggagacacagaaatgagctctcttcaaatca	718
QY	298	-----gcttttctcatagagcatcagaagaagagacttgagcgaattgagctc	342
Db	719	ccctctctctaatagagctctttcttcatagacatcagaagaagagacttgagcgaattgagctc	778
QY	343	aggaatttaaaagagacttgaccataaccgaatcttcttgaaactttatccctctgttgaag	402
Db	779	aggaatttaaaagagacttgaccataaccgaatcttcttgaaactttatccctctgttgaag	838
QY	403	atattgacatatttttaacttgcgaanaagcacccttcacagaaacctgctcgtctgtagagaag	462
Db	839	atattgacatatttttaacttgcgaanaagcacccttcacagaaacctgctcgtctgtagagaag	898
	463	accaaaaccatacccgctgagacgaactgacaccttgaaatgagctctctgcgaagctcttcaga	522
Db	899	accaaaaccatacccgctgagacgaactgacaccttgaaatgagctctctgcgaagctcttcaga	958
QY	523	gcccttcacatcatgtaaggggaaatctgcgcgaagaagcgaagtcacactcgtcgcgcgaacatg	582
Db	959	gcccttcacatcatgtaaggggaaatctgcgcgaagaagcgaagtcacactcgtcgcgcgaacatg	1018
QY	583	ccatgctctcgtgggtcccggaagctgcgaagctctgcgaacagtccaattatgctctctctcc	642
Db	1019	ccatgctctcgtgggtcccggaagctgcgaagctctgcgaacagtccaattatgctctctctcc	1078
QY	643	tcgcgtcccaagagcccaaggctggaacttttgaaacctctctgtaataacatccctggata	702
Db	1079	tcgcgtcccaagagcccaaggctggaacttttgaaacctctctgtaataacatccctggata	1138
QY	703	taactgcacaaatcagaagaacagacatcatgcccacatgctgctcgaagctgcgcgcgaag	762
Db	1139	taactgcacaaatcagaagaacagacatcatgcccacatgctgctcgaagctgcgcgcgaag	1198
QY	763	ttctttctctcttgatgctacatgaaatccaagccccagaactgcgccagaatcagaag	822
	1199	ttctttctctcttgatgctacatgaaatccaagccccagaactgcgccagaatcagaag	1258

[illegible]

Db 2339 catgggaaagcgtcgagaagacacagctgtaatccacatcgaaagagccccaagacct 2398
QY 1963 aacttcccagcaagcgtgtatcttcttcttcaactggaagcaggaatttaagacctg 2022
Db 2399 aacttcccagcaagcgtgtatcttcttcttcaactggaagcaggaatttaagacctg 2458
QY 2023 aggtcacaccccggaatttcagcaattgataagcaagatcacatctctgggaaaa 2082
Db 2459 aggtcacaccccggaatttcagcaattgataagcaagatcacatctctgggaaaa 2518
QY 2083 tattcagctctgcacaaagcctcagcgtgcaaa taaagagatg tgcgtgtgtgctggaa 2142
Db 2519 tattcagctctgcacaaagcctcagcgtgcaaa taaagagatg tgcgtgtgtgctggaa 2578
QY 2143 gctcagcttctgctcctcagcactctgtaagaacttattctcctcagtggtgaaagccagtc 2202
Db 2579 gctcagcttctgctcctcagcactctgtaagaacttattctcctcagtggtgaaagccagtc 2638
QY 2203 ccttcacatagagaagtgaagagcacatcacatctgtaaaaacctgaaaccttgagta 2262
Db 2639 ccttcacatagagaagtgaagagcacatcacatctgtaaaaacctgaaaccttgagta 2698
QY 2263 ttcatgacctacagaa tcaacggtctgcgggtgtgtctgactgacagcttggtgtaacttga 2322
Db 2699 ttcatgacctacagaa tcaacggtctgcgggtgtgtctgactgacagcttggtgtaacttga 2758
QY 2323 agaaccttacaaagctcataatgagtaaacataaagaatgaatgaagaagatgctataaac 2382
Db 2759 agaaccttacaaagctcataatgagtaaacataaagaatgaatgaagaagatgctataaac 2818
QY 2383 tagctgaagcgtgtaaaaacctgaagaagaatggttatttcaattgacccacttctgt 2442
Db 2819 tagctgaagcgtgtaaaaacctgaagaagaatggttatttcaattgacccacttctgt 2878
QY 2443 acattggagaggaatggaatgataatgataagctcgtcgaag tgaacctgtgactctg 2502
Db 2879 acattggagaggaatggaatgataatgataagctcgtcgaag tgaacctgtgactctg 2938
QY 2503 aagaatattcaattgactctcgtctgtctgtctgtctgtcgaatgcaagtgaaaaatctcgtcaga 2562
Db 2939 aagaatattcaattgactctcgtctgtctgtctgtcgaatgcaagtgaaaaatctcgtcaga 2998
QY 2563 attctacaattggtcacaactgagcattcttgattatagaaaaatttaccttgaaaaag 2622
Db 2999 attctacaattggtcacaactgagcattcttgattatagaaaaatttaccttgaaaaag 3058
QY 2623 atggaatgaaagctctctcaactgacacagagatgaacgtgctagaaacagctcaacg 2682
Db 3059 atggaatgaaagctctctcaactgacacagagatgaacgtgctagaaacagctcaacg 3118
QY 2683 caactgactgtccctgtggcgtgtgacgttgcagaagcagcctgagcagcgtgtgaaaaact 2742
Db 3119 caactgactgtccctgtggcgtgtgacgttgcagaagcagcctgagcagcgtgtgaaaaact 3178
QY 2743 tggaggaaggtcccaacactgtaacgttgggttgaaaaactggaactacacagatcacg 2802
Db 3179 tggaggaaggtcccaacactgtaacgttgggttgaaaaactggaactacacagatcacg 3238
QY 2803 agattggaattttaggtgtcaatttttggaaaagaacctctgaaaaaacttccacagattga 2862
Db 3239 agattggaattttaggtgtcaatttttggaaaagaacctctgaaaaaacttccacagattga 3298
QY 2863 atttgcgggaaatcg tgtgagcagtgatgagatgctgtccttcacatggtgtatttgaga 2922
Db 3299 atttgcgggaaatcg tgtgagcagtgatgagatgctgtccttcacatggtgtatttgaga 3358
QY 2923 atttgaagcaattagtggtttttagactttagtaagaagatttaccctgacagcagat 2982
Db 3359 atttgaagcaattagtggtttttagactttagtaagaagatttaccctgacagcagat 3418
QY 2983 tagtcgaagaaacttagcgaagtgtatccaagttaacttctgcgaagaagcagcttg 3042

Db 3419 tagtcgaagaaacttagcgaagtgtatccaagttaacttctgcgaagaagcagcttg 3478
QY 3043 ttgggtggcaatttgatgatgatgatctcagtttatttaagtg 3087
Db 3479 ttgggtggcaatttgatgatgatgatctcagtttatttaagtg 3523

RESULT 4
AAH99581
ID AAH99581 standard; cDNA, 2950 BP.
XX
AC AAH99581;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:416.
XX
KW Human; cancer; HIV infection; human immunodeficiency virus;
KW anti-inflammatory; antirheumatic; antirheumatic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimitagen; cardiovascular; antinaemic; anaemia;
KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; cytostatic;
KW dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000MO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
PI
XX
DR WPI: 2001-457603/49.
DR P-PSDB; AAM25640.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS Claim 1: Page 511-512; 1217pp; English.
XX
CC AAH99166 to AAH9904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antirheumatic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimitagen;
CC cardiovascular; antinaemic; antiagregant; haemostatic; vulnerary;
CC antiulcer; osteopathic; dermatological; antiallergic; antisthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production, or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

XX Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other.

Query Match 62.1%; Score 1946.6; DB 22; Length 2950;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
OY 1181 tgcagtgactcattcctcgagagccttggaccactgtggagacctagctctcggaggtgtgtt 1240
Db 1 tgcagtgactcattcctcgagagccttggaccactgtggagacctagctctcggaggtgtgtt 60

OY 1241 ctcccaagaatttgattctcgactcgaagatgtgtccagcgtgtgaatggatgtctcct 1300
Db 61 ctcccaagaatttgattctcgactcgaagatgtgtccagcgtgtgaatggatgtctcct 120

OY 1301 gacaactggtcctcctgttaataatacagctcaaaaggttcaagccaagataattctt 1360
Db 121 gacaactggtcctcctgttaataatacagctcaaaaggttcaagccaagataattctt 180

OY 1361 tcaagaatcattcagaagagtaacaagcaggaagcagactcagcagtttattgacgtctca 1420
Db 181 tcaagaatcattcagaagagtaacaagcaggaagcagactcagcagtttattgacgtctca 240

OY 1421 tgaagcaagaagagtgagcaagaggaatgttacttgcagaagaatgtgttccatttgcga 1480
Db 241 tgaagcaagaagagtgagcaagaggaatgttacttgcagaagaatgtgttccatttgcga 300

OY 1481 cattacatccacttaagacagcctgtctccggtlacacctgtgtgtatctgttggaaagcac 1540
Db 301 cattacatccacttaagacagcctgtctccggtlacacctgtgtgtatctgttggaaagcac 360

OY 1541 cagggtctgtatgaaagcctcgcgacagagtatatacaacggtcgtcttcgcgaacttc 1600
Db 361 cagggtctgtatgaaagcctcgcgacagagtatatacaacggtcgtcttcgcgaacttc 420

OY 1601 catgcgcaagagagcctctcttggaagcagaatcttgcagaagtgtgaaanaaacacactga 1660
Db 421 catgcgcaagagagcctctcttggaagcagaatcttgcagaagtgtgaaanaaacacactga 480

OY 1661 gcaagaanaattctgaagaagcacaataacatcattccttctgtagagtgtgcatcattata 1720
Db 481 gcaagaanaattctgaagaagcacaataacatcattccttctgtagagtgtgcatcattata 540

OY 1721 tcaagaagatcattccaaatacagccctgtgagccaagaatttgaagcttcttcaagtaa 1780
Db 541 tcaagaagatcattccaaatacagccctgtgagccaagaatttgaagcttcttcaagtaa 600

OY 1781 aagcttatatactaactaaggaagaaatcccgattacttatttgactcttcttgaacattt 1840
Db 601 aagcttatatactaactaaggaagaaatcccgattacttatttgactcttcttgaacattt 660

OY 1841 gcccaattgtgcaaggtgtctcggactcatlaaacttgacttttaaggggagcattgtgc 1900
Db 661 gcccaattgtgcaaggtgtctcggactcatlaaacttgacttttaaggggagcattgtgc 720

OY 1901 ttcatgtggaagaaagcgtcagagaagacaggttggaaatccacatggaagagccccaagaa 1960
Db 721 ttcatgtggaagaaagcgtcagagaagacaggttggaaatccacatggaagagccccaagaa 780

OY 1961 ctacattcccgagggaggtgtatcttcttcttcaactggagagcaggaattcaggaactc 2020
Db 781 ctacattcccgagggaggtgtatcttcttcttcaactggagagcaggaattcaggaactc 840

OY 2021 ggaaggtcacactccggagtttcagcaagtgtgaataagcaaatatcatatctcgtggaa 2080
Db 841 ggaaggtcacactccggagtttcagcaagtgtgaataagcaaatatcatatctcgtggaa 900
```

```
OY 2081 aatatcagctctgcacaaagcctcagagctgtcgaataaagagatgtgtgtgtcgtcg 2140
Db 901 aatatcagctctgcacaaagcctcagagctgtcgaataaagagatgtgtgtgtcgtcg 960

OY 2141 aagcttcagttgtgtcctcagacactgtgaagaacttattctcactgtgtggaagcag 2200
Db 961 aagcttcagttgtgtcctcagacactgtgaagaacttattctcactgtgtggaagcag 1020

OY 2201 tccctccacataagaagtgaagaagcacatcacatctgttaacaaacctgaaacctgag 2260
Db 1021 tccctccacataagaagtgaagaagcacatcacatctgttaacaaacctgaaacctgag 1080

OY 2261 tattcaagacctacagaatcaacagcgtgcgggtgtgtctgtactgacagcttggtaactc 2320
Db 1081 tattcaagacctacagaatcaacagcgtgcgggtgtgtctgtactgacagcttggtaactc 1140

OY 2321 gaagaaccttacaagcctcacaatgtgaataaataaagatgaaatgaaagtgtcattaaa 2380
Db 1141 gaagaaccttacaagcctcacaatgtgaataaataaagatgaaatgaaagtgtcattaaa 1200

OY 2381 actagctgaagcctgtaaaaacctgaagaagatgtgttatttcaatttgaccactgtc 2440
Db 1201 actagctgaagcctgtaaaaacctgaagaagatgtgttatttcaatttgaccactgtc 1260

OY 2441 tgaacttggagaggaatgtgattacatagtcgaagtcctgtcgaagtgaacctgtgacct 2500
Db 1261 tgaacttggagaggaatgtgattacatagtcgaagtcctgtcgaagtgaacctgtgacct 1320

OY 2501 tgaagaanaattcaatagctcctgtctgtctgtctgtcgaatgaggaataatcctgacca 2560
Db 1321 tgaagaanaattcaatagctcctgtctgtctgtctgtcgaatgaggaataatcctgacca 1380

OY 2561 gaattctacaatttgttccaacttgaagcattcttgaattatcaagaanaattaccttgaaaa 2620
Db 1381 gaattctacaatttgttccaacttgaagcattcttgaattatcaagaanaattaccttgaaaa 1440

OY 2621 agatgtgaataatgaagcctcttcataagatcgtacgaagatgaagtgctcagaacagctac 2680
Db 1441 agatgtgaataatgaagcctcttcataagatcgtacgaagatgaagtgctcagaacagctac 1500

OY 2681 cgcactgtatgtgcctcctgtgtgagcgtgcaagcagcctggaagcagcctgttgaaaca 2740
Db 1501 cgcactgtatgtgcctcctgtgtgagcgtgcaagcagcctggaagcagcctgttgaaaca 1560

OY 2741 ttgtgaggaaggtccccaacactcgtlcaagcttgggttgaanaacttgaagactcacagatac 2800
Db 1561 ttgtgaggaaggtccccaacactcgtlcaagcttgggttgaanaacttgaagactcacagatac 1620

OY 2801 agagattgaattttaggttgcaatttcttggaaagaaccccttgaaaccttcagagatt 2860
Db 1621 agagattgaattttaggttgcaatttcttggaaagaaccccttgaaaccttcagagatt 1680

OY 2861 gaatttggcgggaatctgtggaagcagtgatggatgggttgccttcagtggttgatttga 2920
Db 1681 gaatttggcgggaatctgtggaagcagtgatggatgggttgccttcagtggttgatttga 1740

OY 2921 gaatttgaagcaattaggttcttctgaactttagtaactaaagaattctcactgatacagc 2980
Db 1741 gaatttgaagcaattaggttcttctgaactttagtaactaaagaattctcactgatacagc 1800

OY 2981 attagtcagaanaacttgaagcagtgattatccaagttaacttctcgaagaagctagact 3040
Db 1801 attagtcagaanaacttgaagcagtgattatccaagttaacttctcgaagaagctagact 1860

OY 3041 tgttggtgtgcgaatttgaatgtatgatacctcagtgatgttacaagtggttttaactagt 3100
Db 1861 tgttggtgtgcgaatttgaatgtatgatacctcagtgatgttacaagtggttttaactagt 1920

OY 3101 aactgttaaataaagtgttactcgaagcagta 3133
Db 1921 aactgttaaataaagtgttactcgaagcagta 1953
```



```
RESULT 5
AAH34171/c
ID AAH34171 standard; cDNA; 2735 BP.
XX
XX AAH34171;
AC
XX 03-SEP-2001 (first entry)
DT
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1253.
DE
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma; ss.
XX
XX Homo sapiens.
OS
XX MO200122920-A2.
PN
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000MO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
XX
XX P-PSDB; AAG74766.
DR
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3017; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing P.
CC Inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 2735 BP; 791 A; 555 C; 512 G; 876 T; 1 other;
SQ
Query Match 18.0%; Score 564.4; DB 22; Length 2735;
Best Local Similarity 99.8%; Pred. No. 8.2e-156;
Matches 565; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Db 2607 ATGCTGCCCTGGGCGTGGACGTGACGACGCTGACGACCTGTTGTAACATTGGAG 2548
Qy 2748 gaggtccacaaactcgtcaagcttgggttgaaaaacttggaaactccacagataagagatt 2807
Db 2547 GAGGTCCACAACTGTCACGCTTGGTTGAAAACTGGAGACTCAGATACAGAGATT 2488
Qy 2808 agaatttaagtgcatcttttggaaagaacctctgaaanaactccacagctgaatttg 2867
Db 2487 AGAATTTAGTGCATTTTGGGAAAGAACCCCTGAAAACTTCACAGATTGAATTTG 2428
Qy 2868 gcgggaaactcgttggacagtgatgagatgcttgccttcacatgagtgatttgaatct 2927
Db 2427 GCGGGAATCGTGTGACGACGATGATGATGCTGCTGCTTCAATGAGTATTTGAACATCTT 2368
Qy 2928 aagaatttagtgcttttttgaatttagtactaaagaatttccactgatacagattgct 2987
Db 2367 AAGCAATTTAGTGTGTTTGGACTTTAGTAAAGAAATTTTACCTGATCCAGCATTTAGTC 2308
Qy 2988 agaaacttaagccaagtgatcccaagttacttcttgcagaagactaggctgttgg 3047
Db 2307 AGAAACTTAGCCCAAGTCTTATCCAGTTAACTTTTGCAGAGAACTAGGCTTTGGG 2248
Qy 3048 tggcaatttgatgatagtatcagatcagtgatattacagtgctttaaactagtaactgct 3107
Db 2247 TGCCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2188
Qy 3108 taataaagtactcgaagccagta 3133
Db 2187 TAAATAAAGTGTACTCGAAGCCAGTA 2162
RESULT 6
AAI14389
ID AAI14389 standard; DNA; 421 BP.
XX
XX AAI14389;
AC
XX 12-OCT-2001 (first entry)
DT
XX
XX Probe #4322 for gene expression analysis in human cervical cell sample.
DE
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KM cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157278-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001MO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID NO 4322; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC
```

CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcr_sequences.

XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.7%; Score 242; DB 22; Length 421;
 Best Local Similarity 100.0%; Pred. No. 4.4e-61;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2816 aggtgcatcttttggaaagaaaccccttgaaacattccagcagtgtaatttgcgggaaa 2875
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 180 aggtgcatcttttggaaagaaaccccttgaaacattccagcagtgtaatttgcgggaaa 239
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2876 tcgtgtgagcagtgatgagtgcttgccttcatagggtgatttgaaatcctaagaact 2935
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 240 tcgtgtgagcagtgatgagtgcttgccttcatagggtgatttgaaatcctaagaact 299
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2936 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaact 2995
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 300 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaact 359
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2996 tagccaagtgatataccaagtaactttctgcaagaagctagctgttgggtggaact 3055
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 360 tagccaagtgatataccaagtaactttctgcaagaagctagctgttgggtggaact 419
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 3056 tg 3057
 ||
 DB 420 tg 421

RESULT 7

AAI35764
 ID AAI35764 standard; DNA; 421 BP.

AC AAI35764;

DT 17-OCT-2001 (first entry)

DE Probe #4450 used to measure gene expression in human placenta sample.

XX Probe: microarray; human; placenta; antenatal diagnosis;
 XX genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI: 2001-488897/53.
 PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -
 XX Claim 25; SEQ ID No 4450; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.

XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match 7.7%; Score 242; DB 22; Length 421;
 Best Local Similarity 100.0%; Pred. No. 4.4e-61;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2816 aggtgcatcttttggaaagaaaccccttgaaacattccagcagtgtaatttgcgggaaa 2875
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 180 aggtgcatcttttggaaagaaaccccttgaaacattccagcagtgtaatttgcgggaaa 239
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2876 tcgtgtgagcagtgatgagtgcttgccttcatagggtgatttgaaatcctaagaact 2935
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 240 tcgtgtgagcagtgatgagtgcttgccttcatagggtgatttgaaatcctaagaact 299
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2936 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaact 2995
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 300 agtgttttttgacttagtaactaaagaatttctactgatacagcattagtcagaaact 359
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2996 tagccaagtgatataccaagtaactttctgcaagaagctagctgttgggtggaact 3055
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 360 tagccaagtgatataccaagtaactttctgcaagaagctagctgttgggtggaact 419
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 3056 tg 3057
 ||
 DB 420 tg 421

RESULT 8

AAI04213
 ID AAI04213 standard; DNA; 421 BP.

AC AAI04213;

DT 09-OCT-2001 (first entry)

DE Probe #4204 used to measure gene expression in human breast sample.

XX Probe: human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.

PF 29-JAN-2001; 2001WO-US00661.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI: 2001-476286/51.

PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID NO 4204; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosticating diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;
Query Match 7.7%; Score 242; DB 22; Length 421;
Best Local Similarity 100.0%; Pred. No. 4.4e-61;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2816 agtgcattttttgaaagacccctgaaacttcacagcagtgtaattggcgga 2875
DB 180 agtgcattttttgaaagacccctgaaacttcacagcagtgtaattggcgga 239
QY 2876 tcgtgtgacagtgatgatgctgcttcacatggtgtatttgaagacttaagcaatt 2935
DB 240 tcgtgtgacagtgatgatgctgcttcacatggtgtatttgaagacttaagcaatt 299
QY 2936 agtgtttttgacttagtactaaagaattctctaccgtatccagcattagtcagaact 2995
DB 300 agtgtttttgacttagtactaaagaattctctaccgtatccagcattagtcagaact 359
QY 2996 tagcgaagtgtatccaaagttaactttctcgaagaactgagctgtgtgtgtaactt 3055
DB 360 tagcgaagtgtatccaaagttaactttctcgaagaactgagctgtgtgtgtaactt 419
QY 3056 tg 3057
DB 420 tg 421
RESULT 9
AA123590
D AA123590 standard; DNA; 220 BP.
XX
AC AA123590;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #13523 for gene expression analysis in human cervical cell sample.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PS
PS Claim 25; SEQ ID NO 13523; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
Query Match 7.0%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 9.1e-55;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2865 ttggcgggaactggtgtgacagtgatgatgctgcttcacagtggtattggaat 2924
DB 1 ttggcgggaactggtgtgacagtgatgatgctgcttcacagtggtattggaat 60
QY 2925 cttaagcattagtggtttttgacttagtactaaagaattctaccgtacagcacta 2984
DB 61 cttaagcattagtggtttttgacttagtactaaagaattctaccgtacagcacta 120
QY 2985 gtcaagaacttaagcaagtggtatccaaagttaactttctgcaagaagctagctgtt 3044
DB 121 gtcaagaacttaagcaagtggtatccaaagttaactttctgcaagaagctagctgtt 180
QY 3045 ggggtgcaattgtagtgatgatcctcagtggttattacag 3084
DB 181 ggggtgcaattgtagtgatgatcctcagtggttattacag 220
RESULT 10
AA148904
ID AA148904 standard; DNA; 220 BP.
XX
AC AA148904;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #17590 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488897/53.
 DR
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 17590; 654bp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
 Query Match 7.0%; Score 220; DB 22; Length 220;
 Best Local Similarity 100.0%; Pred. No. 9.1e-55;
 Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2865 ttggcgggaaatcgtgtgagcagtgatgagctgcttccttaagtggtattggaat 2924
 Db 1 ttggcgggaaatcgtgtgagcagtgatgagctgcttccttaagtggtattggaat 60
 QY 2925 cttaagcaattagtggttttctgacttagtactaaagaatttctactatccagcat 2984
 Db 61 cttaagcaattagtggttttctgacttagtactaaagaatttctactatccagcat 120
 QY 2985 gtcaagaacctagccaagtgtatccaaagtttaactttctgcaagaagctagctgtt 3044
 Db 121 gtcaagaacctagccaagtgtatccaaagtttaactttctgcaagaagctagctgtt 180
 QY 3045 ggggtgcaatttgatgatgatgatccagtggtattacaag 3084
 Db 181 ggggtgcaatttgatgatgatgatccagtggtattacaag 220
 RESULT 11
 AAI09206
 AAI09206 standard; DNA; 220 BP.
 AC AAI09206;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Probe #9197 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 PD
 XX 09-AUG-2001.
 PE 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-476286/51.
 DR
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 9197; 322bp; English.
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
 Query Match 7.0%; Score 220; DB 22; Length 220;
 Best Local Similarity 100.0%; Pred. No. 9.1e-55;
 Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2865 ttggcgggaaatcgtgtgagcagtgatgagctgcttccttaagtggtattggaat 2924
 Db 1 ttggcgggaaatcgtgtgagcagtgatgagctgcttccttaagtggtattggaat 60
 QY 2925 cttaagcaattagtggttttctgacttagtactaaagaatttctactatccagcat 2984
 Db 61 cttaagcaattagtggttttctgacttagtactaaagaatttctactatccagcat 120
 QY 2985 gtcaagaacctagccaagtgtatccaaagtttaactttctgcaagaagctagctgtt 3044
 Db 121 gtcaagaacctagccaagtgtatccaaagtttaactttctgcaagaagctagctgtt 180
 QY 3045 ggggtgcaatttgatgatgatgatccagtggtattacaag 3084
 Db 181 ggggtgcaatttgatgatgatgatccagtggtattacaag 220

RESULT 12
 AAF58252/C
 ID AAF58252 standard; DNA; 936 BP.
 XX
 AC AAF58252;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Oligonucleotide D1835.
 XX
 KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 XX
 PN WO200107665-A2.
 PD
 XX 01-FEB-2001.
 PE 26-JUL-2000; 2000WO-US20476.
 PR 26-JUL-1999; 99US-0145695.

```

PR 17-MAR-2000; 2000US-0190259.
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX
PI Umek RM;
XX
XX WPI: 2001-159728/16.
DR
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
XX
XX Query Match 3.0%; Score 94.8; DB 22; Length 936;
XX Best Local Similarity 0.8%; Pred. No. 2.3e-17;
XX Matches 6; Conservative 461; Mismatches 313; Indels 0; Gaps
XX
XX 2341 taatgataactaaagatgaatgagaagatgcatataaactagctgaagcccgaaaa 2400
XX 787 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2401 accctgaagaagatgcttattattcattgaccacctgctgacattggaagggaatg 2460
XX 727 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2461 attacatagtcgaagctctctgtaagtgaaacctgtgaccttgaaagaattcaatgct 2520
XX 667 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2521 cctgtcgtctgctgctgaatgacgagaaatcctgtcctcagaattccaaattggtca 2580
XX 607 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2581 aactgagacctgtgattatcagagaatctacctggaagaatgaaatgagctctc 2640
XX 547 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2641 atgaactgacgcacagatgaacgctgctagaacagctcaaccgactgctgctgg 2700
XX 487 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2701 gctgtgacgtgcaagcgacgctgagcagcgttgaacatttggagaaggtccacaac 2760
XX 427 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2761 tcgtcaagcttggtctggaanaactgagagactcacagatacagataatgtagtg 2820
XX 367 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2821 catcttttggaaagaacctctgaanaacttcacagcagltgaaattggtcggaatactg 2880
XX 307 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2881 tgaagcagtgatgagcttgccttcacatgagtgcttttgagaatcttaagcaataagtg 2940
XX 247 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 2941 ttttgacttagtactaaagaatttcaaccgactccagcatagtcgaanaacttagcc 3000
XX 187 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
XX
XX 3001 aaggttttccaaagtaacttttcgcaagaagcctgcttctgggttgcaatttgatg 3060

```

[illegible]

```
QY 2581 aactgacatcttattatcagaataattacctggaaaagatggaatgagctctc 2640
Db 547 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488
QY 2641 atgaactgacagacagatgacgctgtagaacagctcacgcgactgagctgccttgg 2700
Db 487 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 428
QY 2701 gctgtgacgtgcaagcagcctgtagacgctgttgaacatttggaggagctccacaac 2760
Db 427 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368
QY 2761 tcgtcaagcttggttgtaaaactggaactcacagatacagagattagaatttagtg 2820
Db 367 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 308
QY 2821 catttttgaagaagacccctcgaaaactccagcagcttgaattggcggaatcg 2880
Db 307 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248
QY 2881 tgaagcagtgatgagctgctcctcatggtgtatttgaagaacttaagcaattagtg 2940
Db 247 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188
QY 2941 ttttgaactttagtactaagaattctacactgataccagcatatgacgaatactagcc 3000
Db 187 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 128
QY 3001 aagtgatccaagttaactttctgcaagaagcaggctgttggcggaatttgatg 3060
Db 127 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68
QY 3061 atgagatcagtgatgtacttaccagtgctttaaactgaactgcttaataaagtgta 3120
Db 67 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 8

RESULT 14
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
DE Oligonucleotide D1954.
XX
OS Electron-transfer group; ETM; mismatch; genotyping;
   gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR MPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
PS Example 6; Page 127; 159pp; English.
```

```
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 3.0%; Score 94.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 2,3e-17;
Matches 6; Conservative 461; Mismatches 313; Indels 0; Gaps 0;

QY 2341 taatgatacacaagaatgaaatgaaagatgataaactagctgaagcctgaaa 2400
Db 787 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728
QY 2401 acctgaagaagatgtgtatttattcatttgaaccactgtctgacattggaaggaatgg 2460
Db 727 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668
QY 2461 attacatagtaagctctcctgaagtgaacctgtgacctgaagaataatcattagct 2520
Db 667 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 608
QY 2521 cctgtgctgtctgcaaatcagtcagtgaaatcctagctcagaatcttcaattgttca 2580
Db 607 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548
QY 2581 aactgacatcttattatcagaataattacctggaaaagatggaatgagctctc 2640
Db 547 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488
QY 2641 atgaactgacagacagatgacgctgtagaacagctcacgcgactgagctgccttgg 2700
Db 487 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 428
QY 2701 gctgtgacgtgcaagcagcctgtagacgctgttgaacatttggaggagctccacaac 2760
Db 427 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368
QY 2761 tcgtcaagcttggttgtaaaactggaactcacagatacagagattagaatttagtg 2820
Db 367 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 308
QY 2821 catttttgaagaagacccctcgaaaactccagcagcttgaattggcggaatcg 2880
Db 307 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248
QY 2881 tgaagcagtgatgagctgctcctcatggtgtatttgaagaacttaagcaattagtg 2940
Db 247 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188
QY 2941 ttttgaactttagtactaagaattctacactgataccagcatatgacgaatactagcc 3000
Db 187 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 128
QY 3001 aagtgatccaagttaactttctgcaagaagcaggctgttggcggaatttgatg 3060
Db 127 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68
QY 3061 atgagatcagtgatgtacttaccagtgctttaaactgaactgcttaataaagtgta 3120
Db 67 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 8

RESULT 15
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
```

XX 24-APR-2001 (first entry)
XX Oligonucleotide D2004.
DE Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX Synthetic.
OS
XX WO200107665-A2.
PN
XX 01-FEB-2001.
PD
XX 26-JUL-2000; 2000MO-US20476.
PF
XX 26-JUL-1999; 990S-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA
XX
XX UmeK RM;
XX
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 128; 159pp; English.
PS
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other:

Query Match 3.0%; Score 94.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 2.3e-17;
Matches 6; Conservative 461; Mismatches 313; Indels 0; Gaps 0;

QY 2341 taatgatacaataaagaatgaagaagaagatctataaactagctgaagcctgaaa 2400
DB 787
Y 2401 acctgaagaagatggtttatttcaattgacccactgctgacattgagaggaaatg 2460
DB 727
QY 2461 atacatagtaagtcctctgtaagtaagcctgtaagcctgaagaatcaattagctc 2520
DB 667
QY 2521 cctgctgctgctgcaaatgcaatgcaaatcctagctcgaatctcaccaattggtca 2580
DB 607
QY 2581 aactgagactcttgattatcagaataattcactggaagaatgaaatgaagccttc 2640
DB 547
QY 2641 atgaactgacgaagaatgaacgtctagaacagctcacgcactgactgctgctg 2700
DB 487
QY 2701 gctgtgacgtgcaagcagcctgagcgcctgttgaaacattggaagaggtcccaaac 2760
DB 427

QY 2761 tcgtcaagcttggttgtaaaaactggaagactcacagatacagaatagattagtg 2820
DB 367
QY 2821 catlttgaaagaaccctctgaaaaactcagcagctgaaattgcggaatcgtg 2880
DB 307
QY 2881 tgaagcagtgatgagctgctgctcctcagtggtgatttgagaatcctaagaatagtg 2940
DB 247
QY 2941 ttltgacttagtaactaaatctcaccatgacacgactagtcagaactagcc 3000
DB 187
QY 3001 aagtgatccaaagttaactttctgcaagaagctgagctgtggtgcaattgag 3060
DB 127
QY 3061 atgatgctcagtggtattacaggtgctttaaactagtaactgcttaataaagtga 3120
DB 67

Search completed: March 25, 2002, 11:44:40
Job time: 3264 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 10:50:15 ; Search time 2764.94 Seconds

(without alignments)
12176.222 Million cell updates/sec

Title: US-09-697-089-1

Sequence: 1 cgcctagcccggtggaag.....aagtgactcgaagccagta 3133

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_hiv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	551	17.6	552	13	AQ309404 CITBI-EI-
C 2	477.4	15.2	480	10	AV719179 AV719179
C 3	407.8	13.0	602	13	AZ360053 1M0103H11
C 4	364	11.6	364	10	AI263294 q57b01.x
C 5	347	11.1	371	10	AV656315 AV656315
C 6	308.8	9.9	476	11	BG210375 RST29913
C 7	282.8	9.0	470	13	AO624020 HS_5378.B
C 8	249	7.9	261	10	AM337918 he12h11.x
C 9	225.2	7.2	404	13	AO889169 HS_2161.B
C 10	170	5.4	353	13	AQ320928 RPTC11-93
C 11	165.8	5.3	840	11	BF207840 601862546
C 12	162	5.2	509	10	AW418826 ha21e11.x

13	157.2	5.0	219	13	AQ283886	AQ283886 RPTC11-78
14	153.8	4.9	251	11	BF903662	BF903662 IL2-MT018
15	146	4.7	630	13	AQ112439	AQ112439 CIT-HSP-2
C 16	124.8	4.0	412	11	H25984	H25984 y156g07.r1
C 17	106	3.4	728	13	AZ720059	AZ720059 RPTC-24-8
C 18	93.4	3.0	499	10	AI023795	AI023795 ox08d03.x
C 19	66.2	2.1	429	13	AZ484615	AZ484615 1M0311P08
C 20	55	1.8	635	13	B58691	B58691 CIT-HSP-201
C 21	49.2	1.6	668	13	AZ762115	AZ762115 1M0556M15
C 22	46.8	1.5	625	13	AZ614134	AZ614134 1M0442N17
C 23	44.8	1.4	546	13	AZ362463	AZ362463 1M0107N03
C 24	42.2	1.3	1023	10	BE306291	BE306291 601103781
C 25	40.8	1.3	835	10	BE643259	BE643259 Cr-12_8_F0
C 26	40.6	1.3	361	10	AM504276	AM504276 UT-HF-BN0
C 27	40.2	1.3	783	11	BG563851	BG563851 602584655
C 28	40	1.3	424	13	AQ638259	AQ638259 927P1-17A
C 29	40	1.3	518	10	AA669206	AA669206 ad93b07.s
C 30	40	1.3	833	13	AZ533736	AZ533736 ENTCH45TF
C 31	39.8	1.3	939	11	BE887277	BE887277 601508592
C 32	39.6	1.3	467	10	AM988289	AM988289 u905f04.Y
C 33	39.4	1.3	363	10	BE544063	BE544063 601069826
C 34	39.4	1.3	523	11	BF259322	BF259322 HYSMEF001
C 35	39.4	1.3	549	11	BG109381	BG109381 602280435
C 36	39.4	1.3	573	11	BF257885	BF257885 HYSMEF001
C 37	39.4	1.3	611	11	BF253881	BF253881 HYSMEF000
C 38	39.4	1.3	754	11	BF258252	BF258252 HYSMEF001
C 39	39.4	1.3	786	11	BF266655	BF266655 HY-CRA001
C 40	39.4	1.3	854	11	BE881131	BE881131 601482001
C 41	39.2	1.3	759	10	AI729528	AI729528 BMLGH1135
C 42	39.2	1.3	806	13	AQ491970	AQ491970 BNL6162.m
C 43	39.2	1.3	1101	13	AL106910	AL106910 Drosophila
C 44	39	1.2	1201	13	CNS007JU	AL067358 Drosophila
C 45	38.6	1.2	574	10	AW658668	AW658668 95275 MAR

ALIGNMENTS

RESULT 1
AQ309404/c 552 bp DNA 22-DEC-1998
LOCUS CITBI-EI.2528J13.rf CITBI-EI Homo sapiens genomic clone 2528J13,
DEFINITION DNA sequence.
ACCESSION AQ309404 GI:4041438
VERSION AQ309404.1
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other-GSSs: CITBI-EI-2528J13.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@ligr.org
Citations are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.ligr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
Location/Qualifiers
1..552
/organism="Homo sapiens"

REFERENCE 1 (bases 1 to 602)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
 and Wright, D., Weiss, R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0103 row: H column: 11
 Seq primer: CGTGTAAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 602.

FEATURES
 source
 1. 602
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCG1M0103H11"
 /clone_1lb="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g1473214[gb|AF129072.1]), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance.

BASE COUNT 139 a 153 c 139 g 171 t
 ORIGIN

Query Match 13.0%; Score 407.8; DB 13; Length 602;
 Best Local Similarity 80.3%; Pred. No. 2e-97;
 Matches 478; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 1171 aagctgtgctgcgaagtgactcttcgagaccctgagcctgtgagactagctcgg 1230
 DB 595 AGGTGGAGGCTTCAGGTGATTTCACAGGAGCTTAGACTGTGGAGACTGGCCCTAG 536
 QY 1231 aggtgtgtctcccaagaattgatctgaactgcagatgtgccagctgaatgag 1290
 DB 535 AAGGTGTGTTCGCCCAAAATTTGATTTTGACCCGAGCATGGGTCCAGCATGAACGAG 476
 QY 1291 atgtctgtgacaaactgcctctctgtaataacagctcaaaagttcaagccaagt 1350
 DB 475 ACGTCTGTGTGACAAATAGGCTCTCTGTAGTACACAGCTCAGAGGCTGAAGCCAGT 416
 QY 1351 ataattcttcacaaagtcattcagagatcacagcagagcagagctcagcgtttat 1410
 DB 415 ATAAATCTTTCATTAATCATTTTCAGAGAGTACAGCGAGGTGGAGACTAGCAGTTTGC 356

QY 1411 tgacctcatgagccagaagagtgacccaagggaaatggtacttgcagaaaatggtt 1470
 DB 355 TGAGCTCAAAAGAGCGCAGAGAGGTGACCAAAAGGAAACAGCTACTTAACAATGCTT 296
 QY 1471 ccaattcgacattacatccattatagacctgtctcggtaaacctgttgatcgt 1530
 DB 295 CCATCTCTGACATCATACATCCCTATATAGCATCTGCTCTACACGTGTGGTGTCCA 236
 QY 1531 tggagaccaccagggctgttatgagacccctcgacagctgtatcaaaagctgtc 1590
 DB 235 CAGAGCAACCAAGGCGGTCTCATGAGGCCTTGCAGATGTTATCAGACGCGACCTTAC 176
 QY 1591 tcgagcttcctatgcgccaagagagccctctgtgagacaagaaatcttgaagttga 1650
 DB 175 AAGGACTTTCAGTACCAAGAGGCTCTCTGTGAGCGAAGATCAATCAGAGTCTGAGAA 116
 QY 1651 aacacctgagcaagaatcttgaagacataacatcaatctcttgttagagttgga 1710
 DB 115 ATACCACGTGACCAAGATGTTCTGAAGCCATCAATGTAATTCCTGCTGAGTGTGCA 56
 QY 1711 tccattatcacagagatcatccaatcagccctgagcgaagttaagc 1765
 DB 55 TCAATTGTCTCAGAGAGTATGCTTAATCAGACCTGAGCCAGAAATTTGAGC 1

RESULT 4
 LOCUS AI263294/c 364 bp mRNA EST 03-FEB-1999
 DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
 mRNA sequence.
 ACCESSION AI263294
 VERSION AI263294.1 GI:3871497
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 364)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 clone distribution: NCI-CGAP clone distribution Information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbtrp/image/image.html
 Insert length: 2146 Std Error: 0.00
 Seq primer: -40UP from G1bco
 High quality sequence stop: 364.

FEATURES
 source
 1. 364
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2005417"
 /clone_1lb="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"

BASE COUNT 117 a 84 c 55 g 108 t
 ORIGIN

Query Match 11.6%; Score 364; DB 10; Length 364;
 Best Local Similarity 100.0%; Pred. No. 8.2e-86;
 Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2766 aagcttggttgaaactggaagctcacagatcacagagatttagtgatctt 2825

```

Db 364 AGCTTGGGTTGAAAAACGAGACTACAGATACAGAGATTAGATTAGTGCATT 305
Oy 2826 ttgggaagaacccctcgaagaaactcagcagctgaatttggcgggaactcgtggaac 2885
Db 304 TTTGGAGAAACCCCTCTGAAAACTTCAGCAGTTGAAATTTGGCGGAATCTGTGACC 245
Oy 2886 agtgaatgagctgtgccttcacatgggtgcatcttgagaacttaagaactagttctt 2945
Db 244 ACTGATGATGGCTTGCCCTTCATGGGTGATTTGAGAACTTAAGCAATTAGGTTT 185
Oy 2946 gactctagctacaaagaattctcactgataccaagcattgtaagaacttagccaagt 3005
Db 184 GACTTTAGTACAAAGAAATTTCTACCTGATCCAGCATTAAGTCAAGAACTTAGCCAAAGT 125
Oy 3006 ttatccaagttaactctctcgaagaagctagagctgtgttggtgtggaatttaagatga 3065
Db 124 TTAATCAATTAATCTTTTTCAGAGAGCTAGGCTTGTTGGGTGCAATTTGATGATGAT 65
Oy 3066 gatctcagctgtattacaaagctgtccttaactagtaactgcttaataaagtgtactga 3125
Db 64 GATCTCAGTGTATTTACAGTGTGCTTTAACTAGTAACTGCTTAATAAAGTGTACTGCA 5
Oy 3126 agcc 3129
Db 4 AGCC 1
RESULT 5
LOCUS AV656315 371 bp mRNA EST 07-SEP-2000
DEFINITION AV656315 GLC Homo sapiens cDNA clone GICEOAL0 3', mRNA sequence.
ACCESSION AV656315
VERSION AV656315.1 GI:9877329
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 371)
Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
Xue,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens cDNA clone
Unpublished (2000)
TITLE Homo sapiens cDNA clone
JOURNAL
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzgchgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GICEOAL0"
/clone_lib="GLC"
/rissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 112 a 85 c 91 g 82 t 1 others
ORIGIN
Query Match 11.1%; Score 347; DB 10; Length 371;
Best Local Similarity 98.6%; Pred. No. 2.8e-81;
Matches 350; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Oy 1908 gaaaagctgcgaagaacacaggtggaatccacatgaaagagccccaagaactacatt 1967
Db 1 GAAAGGCTGCAGAAACACAGAGTGAATCCACATGGAAGAGGCCCAAAACCTACATT 60
Oy 1968 cccagcaggagctgtatccttcttctcactggaagcaggaaattcgaacttgagtc 2027
Db 61 CCCAGCAGGCGCTGATCTTGTCTTCAACTGGAACAGGAATTCAGGACTCTGGAGGTC 120
Oy 2028 acactcgggatttcagcaagtgaatgaagaagatcacatccttgagggaatattc 2087
Db 121 ACACCTCCGGGATTTTCAGCAAGTTGAATTAAGCAGATATATGATTCGGGAAAATATTC 180
Oy 2088 agctctgcacaaagcctcagcgtcgcacataaagagatgctgtgtgtgagcagctc 2147
Db 181 AGCTTGCCACAAAGCCTCAGGCTGCMAATTAAGATATGATGTGTGTGCTGGAAGCCTC 240
Oy 2148 agttgtcctcagcagcctgtgaagaacttattctcattgtgtggaagccagctcc 2207
Db 241 AGTTGTGCTCTCAGCACCTGTAAAGACATTTATCTCTCATGTGTGGAACCCAGTCCCTC 300
Oy 2208 accatagaagatgagagagcacatcacatctgtaacaacactgaaacacttgagta 2262
Db 301 ACCATAGAAAGATGAGAGGACACATCTGTACCAACCTGGACAGCTTGGGTA 355
RESULT 6
LOCUS BG210375 476 bp mRNA EST 21-APR-2001
DEFINITION RST29913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG210375
VERSION BG210375.1 GI:13732062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 476)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velloso,N., Hess,J., Cottrien,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducart,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
JOURNAL
COMMENT Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaln@athersys.com
High quality sequence stop: 360.
FEATURES
source
1..476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 120 a 99 c 107 g 149 t 1 others
ORIGIN
Query Match 9.9%; Score 308.8; DB 11; Length 476;
Best Local Similarity 97.8%; Pred. No. 4.2e-71;
Matches 313; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Oy 2811 atttagtgcatttttggaagaacccctcgaagaaactccagcagttgattgagc 2870

```

Dp	145	ACTATTAAGCATTTCTTGGAAAGAACCCTCGAATAAACCTTCACGCACTTGATTGGCG	204
Qy	2871	ggaatcgtctgagcaagtgtatgatgatgccttcctcatgggttatttgagaacttaag	2930
Dp	205	GGAATTCGTCTGCAGCAGTATGATGATGGCTTGCTGTGGGTATWTTGAAATCTTAAAG	264
Qy	2931	caattaagtgcttttgaccttaagtaactaaagaattctaactgtaccgaacttagtaca	2990
Dp	265	CAATTAGTGTTTTTGACTCTTAGTACTAAGAATTTCTACTGATCCAGCATTTAGTCAGA	324
Qy	2991	aacttaaccaagtgtatacccaagttaactcttcctcgcaagaagctaaggcttgtgggtg	3050
Dp	325	AAACTTACCAAGTGTATTCACAGTTACTTTTCTGCAGAAGAAGCTAGGCTGTGGGTGG	384
Qy	3051	caatttgatgatgatgatcatcctcaagtgtattataccaagtgctttaactagtaactgttaa	3110
Dp	385	CAATTGTATGATGATGATGATCTCAGTCTCTTACAGTGTCTTTAAACTAGTAACTGCTTAA	444
Qy	3111	ataaagtctactcgagacca 3130	
Dp	445	ATTAAGTGTTACTCGAAGCCA 464	
RESULT	7		
LOCUS	A0624020	470 bp DNA GSS	16-JUN-1999
DEFINITION	HS_537B_B2_C12.SP6E RPCR-11 Human Male BAC Library Homo sapiens genomic clone Plate=954 Col=24 Row=F, DNA sequence.		
ACCESSION	A0624020		
VERSION	A0624020.1 GI:5086412		
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 470) Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		
MEDLINE	99380589		
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCR-11. For BAC library availability, please contact Plier de Jong (plierdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (http://www.htsc.washington.edu) http://www.htsc.washington.edu Plate: 954 row: F column: 24 Seq primer: SP6 Class: BAC ends High quality sequence stop: 470. Location/Qualifiers 1..470 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="plate=954 Col=24 Row=F" /clone_lib="RPCR-11 Human Male BAC Library" /sex="male" /note="Vector: pBACE3.6; site_1: EcoRI; site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRII. Methylation. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"		
FEATURES	source		

BASE COUNT	142 a	101 c	107 g	117 t	3 others
ORIGIN					
Query Match		9.0%;	Score 282.8;	DB 13;	Length 470;
Best Local Similarity	84.1%;		Pred. No. 3.5e-64;		
Matches 317;	Conservative	0;	Mismatches 60;	Indels 0;	Gaps 0;
OY 2007	gaatcaggaactctggaaggtcaacactccggagattcagaagatgaaatgaagaagatac	2066			
Db 2	gattttcagagaccttgagaggtcacactccgggatttcacagatgtaaaatgaagatgac	61			
OY 2067	acattctgggggaaaaatattcagctctgcacaaagccccaagccctgcgaataaagaatggt	2126			
Db 62	agattttctgggggaaaaatatttcagctctgcacaaagccctgcgaataaagaagatg	121			
OY 2127	gctggtctggtgctggaagcctcagtttgcttcctcagcacctgtaagaacattatctctc	2186			
Db 122	gctggtggtggtggaagcctcagtttggtggtggtggtggtggtggtggtggtggtggtg	181			
OY 2187	atggtggaagcagcagctccctccatcaatagaagatggaaggaacatcatctgtlaaac	2246			
Db 182	atggggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	241			
OY 2247	ctgaaaccttgagatcatcatcagcctcaagaatgaaggaagcgcgcgcgcgcgcgcgc	2306			
Db 242	ctgaaaccttgagatcatcatcagcctcaagaatgaaggaagcgcgcgcgcgcgcgcgc	301			
OY 2307	agctcggttaacttaagaagcctcaagaagcctcaagaagcctcaagaagcctcaaga	2366			
Db 302	agctcggttggtggttggttggttggttggttggttggttggttggttggttggttggt	361			
OY 2367	gaagatgctataaact 2383				
Db 362	gagagagacttatctct 378				
RESULT 8					
AM337918/c					
LOCUS	AM337918	261 bp	mRNA	EST	31-JAN-2000
DEFINITION	he12h11.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:291883 3',				
ACCESSION	AM337918				
VERSION	AM337918.1	GI:6834544			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 261)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
	Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@dbi.riken.go.jp Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbfp/image/image.html Seq primer: -400p from Gbco High quality sequence stop: 201. Location/Qualifiers 1..261 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:291883" /clone_lib="NCI_CGAP_CML1" /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL				

/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: PBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RCC11 Human Male BAC Library"
BASE COUNT 170 a 107 c 114 g 162 t
ORIGIN

Query Match 5.4%; Score 170; DB 13; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-34;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2650 tcgacagatgaacgtctagacagcaccgactgctgcctggggtgagc 2709
258 TCGACAGATGAGCGCTAGAGACACTCAGCCACTGAGTGCCTGGGCGCTGAGC 199
2710 tcgacagcagcgtgagcagcctgtgaaacattgagagaggtcccaacatcgtaagc 2769
198 TGCACAGCAGCCTGAGCAGCCTGTTGAAACATTGGAGGAGGTCCACAACTGCTCAAGC 139
Db 198 TGCACAGCAGCCTGAGCAGCCTGTTGAAACATTGGAGGAGGTCCACAACTGCTCAAGC 139
OY 2770 ttgggtgaaactgagagacacagatagagattgattagat 2819
138 TTGGGTGAAACTGAGACTCAGATACAGATTTGAAATTTAGGT 89

RESULT 11
BF207840 840 bp mRNA EST 06-NOV-2000
LOCUS BF207840/c 6018623461 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082074 5',
DEFINITION mRNA sequence.
ACCESSION BF207840
VERSION BF207840.1 GI:11101426
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 840)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1995)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Clontech Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM944 row: 1 column: 11
High quality sequence stop: 636.
Location/Qualifiers
1. 840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4082074"
/clone_1ib="NIH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgcctgcgc); Site_2: SfiI (ggccatagcgc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATCTAGAGGCGGAGCGGCGGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT 276 a 143 c 165 g 256 t
ORIGIN

Query Match 5.3%; Score 165.8; DB 11; Length 840;
Best Local Similarity 92.7%; Pred. No. 5.2e-33;
Matches 229; Conservative 0; Mismatches 12; Indels 6; Gaps 5;

OY 2888 tgaatgagctgctgccttcacagtgatgagaaatcctaagaattagttttga 2947
Db 679 TGATGAAATGCTCCCTCATGAGTGTA-TTGAANTCTTAACCAATTAATGTT-TTTGA 622
OY 2948 cttagaactaagaattctaccgacagcattagtcagaagaattagcc-aagtg 3006
Db 621 CTTTACTACTAAAGAAATTTCTACTCATCA-CATTAGTCTGAAACTTAGCCAAAGTGT 563
OY 3007 tatccagaatattcttcagaagaagcagcttggttggtgcaatttgatgagatg 3066
Db 562 TATCCAAAGTTAACTTTTCTGCAAGAGCTAGGCT--CTTGGGTCCTTTGATGATGATG 505
OY 3067 atctcagtgatattacagtgctttaaactagtaactgcttaataagtgactcga 3126
Db 504 ATCTCAGTGTATTACAGGTGCTTTAACTACTACTGCTTAATAATGATGATCGAA 445
OY 3127 gccagta 3133
Db 444 GCCAGTA 438

RESULT 12
AA418826/c 509 bp mRNA EST 09-FEB-2000
LOCUS AA418826 112111 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2874380 3',
DEFINITION mRNA sequence.
ACCESSION AA418826
VERSION AA418826.1 GI:6946758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 509)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 461.
Location/Qualifiers
1. 509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2874380"
/clone_1ib="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid12 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonoids 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."

FEATURES
source

BASE COUNT 161 a 99 c 80 g 169 t
ORIGIN

Query Match 5.2%; Score 162; DB 10; Length 509;
Best Local Similarity 100.0%; Pred. No. 4,9e-33;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2972 tgaatcagcattagtcagaaacttagcgaagtgtatccaaagtaactttctgcaaga 3031
|||||
509 TGATCCACACATTAGTACGAAAACTTAGCCAAAGTGTATCCAAAGTTACTTTCTGCAGAA 450

3032 agctagagctttgttggtgcaatttgatgatgatctcagtgatattacaggtgctt 3091
|||||
449 AGCTAGGCTTTGTGGGTGGCATTGTGATGATGATCTCAGTGTATTACAGGTGCTTT 390

3092 taactagtaactgcttaataaagtgtactcgaagccagta 3133
|||||
389 TAAACTAGTACTGCTTAATAAGTGTACTCGAAGCCAGTA 348

RESULT 13
LOCUS A0283886 219 bp DNA GSS 27-Apr-1999
DEFINITION R0111-78E13.TV R011-11 Homo sapiens genomic clone R011-78E13,
DNA sequence.

ACCESSION A0283886
VERSION A0283886.1 GI:3910204
KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 219)

AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P., and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building

TITLE Unpublished (1998)
JOURNAL Contact: Mark Adams
COMMENT Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: mdadams@igf.org
Clones are derived from the human BAC library R011-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.igf.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
Location/Qualifiers
1..219
/organism="Homo sapiens"
/db_xref="GDB:7529676"
/db_xref="taxon:9606"
/clone_id="R011-78E13"
/clone_lib="R011-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
R0111 Human Male BAC Library"

BASE COUNT 69 a 45 c 46 g 59 t
ORIGIN

Query Match 5.0%; Score 157.2; DB 13; Length 219;
Best Local Similarity 86.1%; Pred. No. 8.1e-31;
Matches 174; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

2134 tggctggaagccttagtctcctcagcactgtaagaacattattctcatggtg 2193

Db 1 TGGCTGGAAGCCTTAGTCTGCTCCTCAGCACCTGTAGAACATTTATCTCATGCTGG 60
|||||

2194 aagccagtcocctcaccatagagaatgagagacatcacatctgtaacaaacctgaana 2253
|||||

Db 61 AAGCCAGTCCCTGACCATAGAGATGAGAGACATCATCTGTATTAACAACTGGAANA 120
|||||

2254 ccttagatattcatgaactacagaaatcaagcgtccggtgtgtctgactgacagcttg 2313
|||||

Db 121 CTTGATGATTCATGATACCTACAGATCAAGCGCTCGGGTATGTGTATATCATGTGG 180
|||||

2314 gtaactgagaagacattacaaa 2335
|||||

Db 181 TGTGCTTTGTCATCTTAATAAAA 202
|||||

RESULT 14
LOCUS BF903662 251 bp mRNA EST 18-Jan-2001
DEFINITION IL2-MT0180-181200-276-F03 MT0180 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF903662
VERSION BF903662.1 GI:12295121
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 251)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.R.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?IL2-IL2&t2=IL2-MT0180-
181200-276-F03&t3=2000-12-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 96.

FEATURES
Location/Qualifiers
1..251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="MT0180"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (O.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 72 a 62 c 56 g 61 t
ORIGIN

Query Match 4.9%; Score 153.8; DB 11; Length 251;
Best Local Similarity 85.5%; Pred. No. 6.6e-30;
Matches 218; Conservative 0; Mismatches 32; Indels 5; Gaps 4;

218 tggctggaagccttagtctcctcagcactgtaagaacattattctcatggtg 2193

Oy 1042 agaccctctcttcttgatcactcttgcaatccagatgggt-gaaagtatgtccac 1100
 Db 1 AGACCCCTCTCTTTGAGGTCAATCTTTGTGCATCTCTATGAGACAGAAAAGGAGTTCCAC 60
 Oy 1101 tctcacacaacaagaagctgtgtccatacctctatgtalcgttgtatcagaanaacaa 1160
 Db 61 TCTCACACACAAACCAACCTGATCCATCACTCTTATGTATCTAAGATACAGAAAAACAA 120
 Oy 1161 cacaacaataaagtgtygtgcctcgaagtacatctatcttgagagccttgaccactgtgtgagc 1220
 Db 121 CACAAACATTAAGGGGTGTCTC--ATGTGACTCATTTGGAGGCTT-GACCACCTGTGG-GAC 176
 Oy 1221 cttagctctctgagagtgtygttctccccaagtttgatcttcgaactcagagatgtgtccagc 1280
 Db 177 CTAGATCTCGGTGGGTGGCTGTGGCCACCAAGTTTGAATTTGCAACTGCAGATGTGCACCCG 236
 Oy 1281 gtgaatgagagatgc 1295
 Db 237 GCGTACTGAAATGTC 251

RESULT	15
AQ112439	
LOCUS	AQ112439 630 bp DNA
DEFINITION	CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA sequence.
ACCESSION	AQ112439
VERSION	AQ112439.1 GI:3484599
KEYWORDS	GSS.
SOURCE	human
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 630)
AUTHORS	Adams, M. D., Rounsley, S. D., Zhao, S., Bass, S., Linher, K., Golden, K.

TITLE	Author	Journal	Year
Use of a random human BAC End Sequence Database for Sequence-Ready

JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2372C1.TF

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .630

BASE COUNT	188 a	126 c	124 g	192 t
ORIGIN				

Query Match	4.7%	Score 146	DB 13	Length 630
Best Local Similarity	100.0%	Pred. No. 9.1e-28		
Matches 146	Conservative 0	Mismatches 0	Indels 0	Gaps 0
OY 2384	agctgaagcgctgaanaaacctgaagaagatgtgttattatcattgaccacctgtctcga			
	2443			

Db	485	AGCTGAAGCCTGAAAAACCTGAGAGAGATGTGTTTATTCATTGTGACCACTGGTCTGA	544
Oy	2444	catctggagaaggaatgattacataagtcgaagttctctcgaatggaacctgtaaccttga	250
Db	545	CATTGGAAAGGGAAATGGATTACAAAGTCAAGTCTCTCTCAAGTGAACCCCTGTGACCTTGA	604
Oy	2504	agaatctcaattagcttcctcgtcgtct	2529
Db	605	AGAAATTCAATTACTCTCTCTCTCTCT	630

Search completed: March 25, 2002, 11:37:02
Job time: 2807 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:45:36 ; Search time 4399.74 Seconds

(without alignments)
11747.436 Million cell updates/sec

Title:

Sequence: 1 US-09-697-089-1
1313
1 cgcctagcccggtgggaag.....aagttactcgagccagta 3133

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Geneml: *
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.om:*
20: em.or:*
21: em.ov:*
22: em.pat:*
23: em.ph:*
24: em.pl:*
25: em.ro:*
26: em.sts:*
27: em.sy:*
28: em.un:*
29: em.vl:*
30: em.htgo.hum:*
31: em.htgo.in:*
32: em.htgo.in:*
33: em.htg.hum:*
34: em.htg.in:*
35: em.htg.rod:*
36: em.htg.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3133	100.0	3133	9	AY032589
2	2946	94.0	3219	9	AY035391
3	2915	93.0	3396	9	AY027787
4	2912	92.9	3581	9	AF376061
5	1946	62.1	160583	2	AC010968
6	1844	58.9	138909	9	CNS01DS3
7	1290	41.2	1355	9	IR2003417
8	836	26.7	1395	9	AY027788
9	485	15.5	768	9	AY027789
10	318	10.2	162692	9	CNS01DS8
11	318	10.2	185281	2	AC011232
12	293	9.4	578	9	AY027790
13	170	5.4	553	11	G55568
14	25	0.8	160583	2	AC010968
15	23	0.7	175152	2	AC074195
16	23	0.7	183556	2	AC019059
17	23	0.7	188459	2	AC090582
18	23	0.7	204487	2	AC090559
19	22	0.7	87834	9	AP002957
20	22	0.7	151088	9	AC020917
21	22	0.7	151366	33	AC007951
22	22	0.7	153733	2	AC011980
23	22	0.7	155531	2	AL592213
24	22	0.7	156230	2	AP002958
25	22	0.7	159946	2	AC079194
26	22	0.7	186413	2	AL161913
27	22	0.7	194533	2	AL592438
28	21	0.7	7595	4	BTCSK35
29	21	0.7	37906	9	AC005199
30	21	0.7	45496	2	AC017948
31	21	0.7	76526	2	AC027055
32	21	0.7	80346	8	ATM4E13
33	21	0.7	84499	8	ATF23E12
34	21	0.7	86710	2	AC004907
35	21	0.7	130117	2	AC079351
36	21	0.7	157308	2	AC024590
37	21	0.7	183839	2	AC013553
38	21	0.7	184864	2	AC080090
39	21	0.7	194874	8	ATCHRIV82
40	21	0.7	195165	2	AC025566
41	21	0.7	195217	8	ATCHRIV83
42	21	0.7	198935	2	AC068573
43	21	0.7	206137	2	AL593857
44	21	0.7	211030	2	AL513468
45	21	0.7	221341	2	AC092992
46	21	0.7	227194	2	AC020727
47	21	0.7	303367	3	AE003538
48	20	0.6	768	9	AY027789
49	20	0.6	865	2	AC078439
50	20	0.6	1092	2	AC047412
51	20	0.6	1198	14	REOS3NSB
52	20	0.6	1688	8	SCYBR141C
53	20	0.6	2833	8	SCYBR142M
54	20	0.6	3709	3	AF116341
55	20	0.6	12595	8	SCIRAI
56	20	0.6	36631	9	HUMCOL7A1X
57	20	0.6	40937	3	CER01E6
58	20	0.6	45027	3	CHRG39N14
59	20	0.6	45459	2	AC006103
60	20	0.6	83969	9	AC005210
61	20	0.6	86155	9	AL159169
62	20	0.6	86719	9	AP000885
63	20	0.6	88326	3	AC005923
64	20	0.6	88839	3	AC084447
65	20	0.6	91733	2	AL391278
66	20	0.6	93409	2	AC073294
67	20	0.6	102185	2	AC083819
68	20	0.6	106601	9	AL357352
69	20	0.6	109810	9	AL139036
70	20	0.6	109810	9	AL139036

C 71	20	0.6	114169	9	AC010902	Homo sapi
C 72	20	0.6	120733	9	AC022124	Homo sapi
C 73	20	0.6	120931	9	AC008390	Homo sapi
C 74	20	0.6	124347	9	AC010072	Homo sapi
C 75	20	0.6	129837	9	AC004829	Homo sapi
C 76	20	0.6	130981	2	AC087702	Typanoso
C 77	20	0.6	132171	3	AC008370	Drosophi
C 78	20	0.6	137955	2	AL359752	Human DNA
C 79	20	0.6	141079	8	AP002868	Oryza sat
C 80	20	0.6	142203	2	AC079521	Mus muscu
C 81	20	0.6	145013	2	AC010435	Homo sapi
C 82	20	0.6	145576	8	AP002541	Oryza sat
C 83	20	0.6	146545	2	AC090564	Homo sapi
C 84	20	0.6	147728	2	AC034113	Homo sapi
C 85	20	0.6	157559	2	AC020344	Drosophi
C 86	20	0.6	158574	2	AC024895	Homo sapi
C 87	20	0.6	159158	9	AL359915	Human DNA
C 88	20	0.6	161590	2	AC092727	Bos tauru
C 89	20	0.6	164785	2	AC027540	Homo sapi
C 90	20	0.6	166953	3	AC093104	Drosophi
C 91	20	0.6	168666	2	AC074324	Homo sapi
C 92	20	0.6	170279	2	AC027053	Homo sapi
C 93	20	0.6	173933	9	AP002022	Homo sapi
C 94	20	0.6	174098	9	AC005737	Homo sapi
C 95	20	0.6	175086	2	AL157889	Homo sapi
C 96	20	0.6	175213	2	CNS07EF6	Homo sapi
C 97	20	0.6	176039	2	AC073065	Homo sapi
C 98	20	0.6	177290	2	AC013536	Homo sapi
C 99	20	0.6	177916	2	AL589863	Homo sapi
C 100	20	0.6	178861	2	AC068231	Homo sapi

ALIGNMENTS

RESULT	1	AY032589	3133 bp	mRNA	PRI	25-May-2001
LOCUS		AY032589				
DEFINITION		Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.				
ACCESSION		AY032589	GI:13899172			
VERSION						
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 3133)				
AUTHORS		Geddes, B. J., Wang, L., Huang, W.-J., Lavellee, M., Manji, G. A., Brown, M., Junman, M., Morganstern, J., Merriam, S., Glucksmann, A., Distefano, P. S. and Berlin, J.				
TITLE		Human CARD12 is a Novel CED4/Apaf-1 Family Member That Induces Apoptosis				
JOURNAL		Biochem. Biophys. Res. Commun. 284 (1), 77-82 (2001)				
PUBMED		11374873				
REFERENCE		2 (bases 1 to 3133)				
AUTHORS		Berlin, J.				
TITLE		Direct Submission				
JOURNAL		Submitted (15-APR-2001) Neurobiology, Millennium Pharmaceuticals Inc., 640 Memorial Drive, Cambridge, MA 02139, USA				
FEATURES		Location/Qualifiers				
SOURCE		1..3133				
CDS		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		36..3110				
		/note="CARD12: CED4/Apaf-1 family member"				
		/codon_start=1				
		/product="caspase recruitment domain protein 12"				
		/protein_id="AAK38730.1"				
		/db_xref="GI:13899173"				
		/translation="MNFIDNSRALIORMGWTIVKQITDLEFVNNVLRREVNIIICE KVEDDAAGIITHMLIKKSGESCNFLKSLKMNPLDODNGSLPOTSGDDDLA ODLKDYHTPSFLNFPYLGEDIIDILFNLSKTFEPVLMRKDDHHRVQDLTLNGLDA				

QY	1	CGCTTACCCGCGTGGAGAGCTTATCCAGACATATTTCTAAGAGACATAGC	60	LQSPCTIEGSGKSKSTLLRIAMWGSCKKALTKRFVFFELRLSRAQGLFETLCD
DB	1	CGCTTACCCGCGTGGAGAGCTTATCCAGACATATTTCTAAGAGACATAGC	60	QLDIDPGRKQTFPMAMLIKROVLELDLDYENFRQONCEPEIALIKENHREKNVI
QY	61	GAGCCCTATTCAAGAAATGGAATGACTGTTATTAACCAATATCAGATGACCTATTG	120	VTTECLRIHROFAGALTAEGWDMTEAQAILEVEVILKEIALEBILQIOKSRCLRL
DB	61	GAGCCCTATTCAAGAAATGGAATGACTGTTATTAACCAATATCAGATGACCTATTG	120	MKTPELVITCAIQMGSESEPHSHQTTLFHFVYVLYLLOKNNKHKGAAASPTSLH
QY	121	TATGGAATGTTCTGAATGCGCAAGAAATCATATTTGCGCAGAGGAGGAGCAG	180	CGDLAFGEVSHKRPDELODYSVNEDEVILTTGLCYLTOORPKPKFKPSQOET
DB	121	TATGGAATGTTCTGAATGCGCAAGAAATCATATTTGCGCAGAGGAGGAGCAG	180	AGRLSLTISHPEEYVTKNGYIQKAVSISDITSTSSLRITGSSVEATRAVMKH
QY	181	ATGCTGCTAGAGGATCATCTCATGATTTTGAAGAGGTTCAAGTCTGTATACCTCT	240	LAAYOHGCLGLSIAKRPYLMROSLDYSKNTTFOELIKANINISVECGHLYOEST
DB	181	ATGCTGCTAGAGGATCATCTCATGATTTTGAAGAGGTTCAAGTCTGTATACCTCT	240	SKSAISOEFAPFGKSLIYNSGNIPIYLFDFEHLNCAASALDFIKLDYFGAMASM
QY	241	TTCTTAAATCCCTTAAGAGAGTGAACATCCTCTATTCAGAGACTTGATGACAAAGTC	300	EKAEDTGGIIMEEAPETIIPSRVSLFPMKQEPRLLEVLRDPSLANKODITYLCK
DB	241	TTCTTAAATCCCTTAAGAGAGTGAACATCCTCTATTCAGAGACTTGATGACAAAGTC	300	IFSSATSLRIQKRCAGAGSLSLVSTCKNIYSIAYEASPLTIEDRHLITSVNLKT
QY	301	TTTTCATCAGACATCAAGAGAGCTTGAGAGATTTGGCTCAGAGATTAAAGGACTTGT	360	LSIHDLQNRPGGLTDSIGMLKTLFTLMDNITKMNEDATIKLAGLKLKMLSPHL
DB	301	TTTTCATCAGACATCAAGAGAGCTTGAGAGATTTGGCTCAGAGATTAAAGGACTTGT	360	THSIDISGRNMYIYKSSISPEPCDEIEFQIYSCCSANAYVITLQNLNLYKSLTIDS
QY	361	ACCATACCCCATCTTTCTGAACTTTATCCCTTGCGTGAAGATTTATTCATTTATTTTA	420	ENVLEKQNEALHEDIDRMNVLEDTLMLPWGCDVGSUSLILHLEVPQVLKL
DB	361	ACCATACCCCATCTTTCTGAACTTTATCCCTTGCGTGAAGATTTATTCATTTATTTTA	420	KMRLTDTETIRILGAFGKNPKNFQDLNAGRVSDSGWLAFFGVEENLKQVFPDF
QY	421	ACTTGAAGAGACCTTCACAGACCTGTCGTGAGAGAGAGACCAACCATCAGCCGC	480	STKEFLDPALVRLKSQVLSKTFILOEARLVQMOFDDDLISVITGAFKLVYA"
DB	421	ACTTGAAGAGACCTTCACAGACCTGTCGTGAGAGAGAGACCAACCATCAGCCGC	480	36..299
QY	481	TGGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	540	/note="CARD: Region: caspase recruitment domain"
DB	481	TGGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	540	522..1406
QY	541	GGAATCTGGCAAGGCAAGTCCACTGCTGCAAGGCAATGCTGAGGAGGAGGAGGAGG	600	/note="NBS: Region: nucleotide-binding site"
DB	541	GGAATCTGGCAAGGCAAGTCCACTGCTGCAAGGCAATGCTGAGGAGGAGGAGGAGG	600	2001..3107
QY	601	GAAATGCAAGGCTTGACCAAGTTCAATTCGTTCTTCTCGCTGAGGAGGAGGAGG	660	/note="LRR: Region: leucine-rich repeats"
DB	601	GAAATGCAAGGCTTGACCAAGTTCAATTCGTTCTTCTCGCTGAGGAGGAGGAGG	660	BASE COUNT 903 a 691 c 729 g 810 t
				ORIGIN
				Query Match 100.0%; Score 3133; DB 9; Length 3133;
				Best Local Similarity 100.0%; Pred. No. 0;
				Matches 3133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 661 aggtgagcttttgaaccctctgtatcaactccctgataatactctgagcaatcaaga 720
|||||
Db 661 AGGTTGACTTTTGAACCCCTCTGTATCACTCTGGATATACCTGGACATACAGCA 720
QY 721 agcagacatcatgagcctatgctgtaagctgagcagagaggtcttctctctgag 780
|||||
Db 721 AGCAGACATTCATGAGCCTGCTCTGAGAGCTGGCAGAGAGGTTCTTCTTCTTGATG 780
QY 781 gctacacatgaattcaagcccccaagaactgcccagaataatcgagccctgataaagaacc 840
|||||
Db 781 GCTACATATGATTAATCAAGCCCCCAAGACTGCCCAATAATCGAAGCCCTGATTAAGGAATCC 840
QY 841 accgcttcaagaacatgctatcatgctacacactcactgagtgctggaagcacaatacgc 900
|||||
Db 841 ACCGCTTCAAGAACATGCTCATGCTACCACTACATGATGCTCTGAGGCAATACGCTG 900
QY 901 agtttggtccctgactgctgtaagtgaggagata tgaacagaacagcgccccagctctca 960
|||||
Db 901 AGTTTGTCCTGACTGCTGAGGTGGGGATATGACAGAAACAGGCGCCAGGCTCTCA 960
QY 961 tcgagaagtgctgatacaagagagctgctggaagctgtgtctcccaaatcaagaatcca 1020
|||||
Db 961 TCCGAAAGTGCTGATCAAGAGACTTGTGAAGGCTTGTGCTCAATTCAGAAATCCA 1020
QY 1021 gctgcttgaggaatctcatgaagaccctctctctgtgtcatcaactgtgcaatccaga 1080
|||||
Db 1021 GCTGCTTGAGGAATCTCATGAACACCCCTCTCTTGTGTGCTCATCTTGTCATCCAGA 1080
QY 1081 tgggtgaaagtgaattccactcctacacacacacaaacgctgtctcatcactctatgac 1140
|||||
Db 1081 TGGGTAAAGTGAGTTCACACTCTCACACACAAACAGGCTGTCCATCTCTATGATC 1140
QY 1141 tgtgtgtaacgaanaaacaacacacataaagtggtgtgctgaagtgaattcatcgcga 1200
|||||
Db 1141 TGTGTATACGAATAACAAACACAAACATTAAGGTGTGCTGCTCAAGTCTTATTCGGA 1200
QY 1201 gctctgacacactgtgagacactagctctgtagaggtgtgtctcccaagaattgattctg 1260
|||||
Db 1201 GCTGTAACCACTGTGGAGACCTAGCTCTGAGAGGTGTGTCTCTCCACAACTTGTGATTCG 1260
QY 1261 aactgacagatgtgtccagcgtaattgaagatgtccctgtgacaaactggctcctgta 1320
|||||
Db 1261 AACTGACAGATGTGTCCAGGCTGAATGAGGATGTCCGCTGACAACTGGGCTCTCTGTA 1320
QY 1321 aatatcagctcaagaagttcaagcaagaatataattcttcaagaacttcaagaact 1380
|||||
Db 1321 AATATACGCTCAAGGTTCAAGCCAAAGTATTAATTTCTTCAAACTCAATTCAGAGAT 1380
QY 1381 acacagacagacgaagactcaagcagtttatgtacgtctca tgaagcagagaggtgacca 1440
|||||
Db 1381 ACACAGCAGACGACAGACTACAGAGTTTATGTACGCTCATGAGCCAGAGAGGTGACCA 1440
QY 1441 aggggaatgttactgtcagaaaaa tggtttcaattcgaacattacatcaactatagca 1500
|||||
Db 1441 AGGGGAATGTTACTTGTGCAAAAAATGTTCCATTTGCGACATTAATCACTTATATGCA 1500
QY 1501 gctgtgctcggtacacactgtgggtcactgtgtgaagccacacagagctgtatataagcac 1560
|||||
Db 1501 GCTGTGCTCGGTACCTGTGGGTCACTGTGTGAAGCCACAGGCGCTGTATGAAGACC 1560
QY 1561 tcgacagcagtgatcaacaacgctgctctctcgaacttccatcgccaagaagcctctc 1620
|||||
Db 1561 TCGCAGCAGTGTATCAACACGCTGCTTGTGAGCTTTCATCGCCANAGAGCCCTCTCT 1620
QY 1621 gtagacagagaaatcttgcagaaatgtgaaaaaacacacacacagcagaaatctgaaagca 1680
|||||
Db 1621 GGAGACAGGATCTTGTGCAAAAGTGTGAAAAACCACTGAGCAAGAAATCTGAAAAAGCA 1680
QY 1681 taaacatcaattctcttctgtgagatgtgagcatcattatatacaagagagaaatccaat 1740
|||||
Db 1681 TAAACATCAATTCCTTGTGATGAGTGTGCAATTCATTAATTAAGAGATTAATTCATAAT 1740
QY 1741 cagccctgagccaagaattgaaacttcttcaaggtaaagcttatatacaactcag 1800
|||||
Db 1741 CAGCCCTGAGCCCAAGATTTGAAGCTTCTTTAAGGTAAAGCTTATATATCACTCAG 1800
QY 1801 ggaacatcccccgaattactatcttgaactcttgaacattgcccgaattgtcagtgctc 1860
|||||
Db 1801 GGAACATCCCCGAAATCTTATTTGACTTCTTGTGAACATTTGCCCAATTTGTGAAGTCTC 1860
QY 1861 tgaactcatataactgagacttttatagggagactataggtcttcaatgagaaagctgag 1920
|||||
Db 1861 TGAATCTATTAACCTGAACTTTATGAGGGAGCTTAAGGCTTATGAGGAAAGGCTTCAG 1920
QY 1921 aagacagagtggaatccatggaagaagccccagaacactacattccccagagagctg 1980
|||||
Db 1921 AAGACAGAGTGAATCCATGGAAGAGGCCCCAGAAACCTTACATTTCCAGCAGGCTG 1980
QY 1981 tatcttgtcttctcaactgtaagcaaggaattcaggaactctggaagtacacatccggagt 2040
|||||
Db 1981 TATCTTGTCTTCACTGAAGAGAGAAATTCAGAGACTCTGAGAGTCACTCCGGGATTT 2040
QY 2041 tcagcaagttggaatgaagataatacatatctggggaaataatlaagctctgcccaga 2100
|||||
Db 2041 TCAGCAAGTTGAATTAACCAAGATATCAATATCTGGGAAATATTCAGCTCTGCCCAA 2100
QY 2101 gctcgaagctgcaataaagagatgtgctgtgtgtgagcagccctcagctctgctccta 2160
|||||
Db 2101 GCCTCAGGCTGCAAAATTAAGAGATGTCTGTGTGTGCTGGAAGCTTCAGTTGGTCTCA 2160
QY 2161 gaacctgtaagaacattatctctcaatgtgtggaagcagtcctccatcagataagaatg 2220
|||||
Db 2161 GCACCTGTAAAGCAATTAATTCCTCATGTGTGGAAGCCAGTCCCTCCATGAAATG 2220
QY 2221 aagagcaatcatatctgtatacaaaacctgaaacctggaattatcagactcagatc 2280
|||||
Db 2221 AGAGGCAATCACTCTGTAAACAACTGAAACCTTGAGATTCATGACCTACAGATC 2280
QY 2281 aacgctgcccgggtgtgtctgactgacagactgtggttaactgaaagaacttacaagaactca 2340
|||||
Db 2281 AACGCTGCCGGGTGTGTCTGACTGACAGCTTGGGTAACTTGAAGAACCTTACAAAGCTCA 2340
QY 2341 taatgataacataaagaatgaatgaagaaga tgcctataaactagctgaagcctgaaaa 2400
|||||
Db 2341 TAAATGATTAACATTAAGATTAATGAAGAGATCTTAATAAATAGCTGAAGGCTGTGAAA 2400
QY 2401 aactgaagaagaatgtgtttatctcatatgagccactgtctgacattggaagaagatg 2460
|||||
Db 2401 AACTGAAGAAGATGTGTATTCATATTGACCACTGTGTGACATTGTGAGAGAGGAAATGG 2460
QY 2461 attacatagtcagactctgtcaagtgaaactgtgacacttgaagaatcaatcagact 2520
|||||
Db 2461 ATTACATAGTCAAGTCTCTCTCAAGTGAACCTGTGACCTTGAAGAAATTCATTAATCT 2520
QY 2521 cctgtgctgtgtctgcaaatgacagtgaaatcctagctcagaatctcacaattgtgta 2580
|||||
Db 2521 CCTGTGCTGTGTGTGCAAAAGCAGTGAATAATCTCAGCTCGAATCTTCACAATTTGGTCA 2580
QY 2581 aactgaacactctgtattttatcagaanaataactctggaagaagaatggaagaagctctc 2640
|||||
Db 2581 AACTGACATTTCTTGATTTATCAGAAATTAATCTGGAAGAAAGATGGAAGAAAGCTCTTC 2640
QY 2641 atgaactgatacgaacagatgaacgtgtcagaacagctcaccagcactgatagtcctccctgg 2700
|||||
Db 2641 ATGAACATGATCGAAGAGATGAAGTGTGTAAGACAGCTCACCGCAGCTGATGCTCGCTGG 2700
QY 2701 gctgtgagctgcaagcagcctgagcagcctgttgaacattggaagaggtccacacac 2760
|||||
Db 2701 GCTGTGACGTGCAAGGAGGCTGAGCAGCTGTGTGAATAATTTGTGAGAGAGAGTCCCAAC 2760
QY 2761 tcgtcaagcttggtgtgaaanaactggagactcagataaagaagatgaagattgaagtg 2820
|||||
Db 2761 TCGTCAAGCTTGGGTGAATAAATCGAGACTCAACAGATCAAGAGATTAATTAAGTGTG 2820
QY 2821 catcttctgaaagaacccctcgaanaactcgaagcagctggaattgtgaggaagaaatcgt 2880
|||||

Db 2821 CATTTTTGGAAAGAACCTCTGTAACCAAGCTTGAATTTGGCGGGAATCTGTG 2880
 QY 2881 tgaagcagatgatgagcttgcttcacatgagtgatattgagaactcttaagcaattatgt 2940
 Db 2881 tgaagcagatgatgagcttgcttcacatgagtgatattgagaactcttaagcaattatgt 2940
 QY 2941 ttttgaacttgatgactaaagaattctcactgacatcagcatagtcagaaactatgac 3000
 Db 2941 ttttgaacttgatgactaaagaattctcactgacatcagcatagtcagaaactatgac 3000
 QY 3001 aagtgatccaaagttaactttctgcagaagaagctagctgttgggtgcgaattatgag 3060
 Db 3001 aagtgatccaaagttaactttctgcagaagaagctagctgttgggtgcgaattatgag 3060
 QY 3061 atgagatcagtgatgacttaacagtgcttcaactagtaactgcttaataaagtga 3120
 Db 3061 atgagatcagtgatgacttaacagtgcttcaactagtaactgcttaataaagtga 3120
 QY 3121 ctggaagccagta 3133
 Db 3121 ctggaagccagta 3133
 QY 3121 ctggaagccagta 3133
 Db 3121 ctggaagccagta 3133

RESULT 2
 LOCUS AY035391 3219 bp mRNA PRI 24-JUL-2001
 DEFINITION Homo sapiens ICE-protease activating factor mRNA, complete cds.
 ACCESSION AY035391
 VERSION AY035391.1 GI:14334214
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3219)
 Poyet,J.L., Srinivasula,S.M., Thnani,M., Razmarz,M.,
 Fernandes-Alnemri,T. and Alnemri,E.S.
 Identification of Ipaf, a human caspase-1-activating protein
 related to Apaf-1
 J. Biol. Chem. 276 (30), 28309-28313 (2001)
 2 (bases 1 to 3219)
 Poyet,J.L., Srinivasula,S.M., Fernandes-Alnemri,T. and
 Alnemri,E.S.
 Direct Submision
 Submitted (16-MAY-2001) Microbiology and Immunology, Thomas
 Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107,
 USA

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

FEATURES
 Source
 CDS
 1..3219
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="821-22"
 145..3219
 /note="IPAF: CED4/Apaf-1 family member; caspase-associated
 recruitment domain containing protein"
 /codon_start=1
 /product="ICE-protease activating factor"
 /protein_id="AAK59843.1"
 /db_xref="GI:14334215"
 /translation="MNFTRKNSRALIORMGMTVTKQITDDLFVNVNLRBENVNICE
 KVEDDAAGIIHMLKSGESCNFLKLMWNYPLFQDLNGOSLPHQTSGLDDILA
 ODLDYHTPFLENYPLGEDIDILFNKSTPEVPLMRKQONHRRVQLNLNGILOA
 LSPCITGESGKSKSTLRIAMIMSGSKALFKFVPLRSLRNOGSLPFTICD
 OLDIPGTRIKQTFMAMLIKRLQRYPLLDGYNEKRPONCEPEIALIKENIRKMWI
 VTTTECLRHTRFGALTAEGVDMTEAOLIREVILKELAEGLLLQIQSRCLNLI
 MKPLFVVITCAIOMGESEFHSHTQTLFHTFYDLLQKNKHKKHGAASPLIRSLD
 CGDLAEGVFSEHKPELODVSVMEDVLTGLCKYTAQRFKRYKFFPKSROEY
 AGRRSLSTLSEPEVTKGNGYLOKNVSIISDITSTYSLLRYTGGSSVEMRAVMKH
 LAAYVHGCLIGLSTAKRPLROESIOSVKNTTDEILKAININSEVCKGHVQEST
 SKSALSDSEFEAFPOGKSLYINSGNIPDLFDFEHLPCASALDEITKIDFYGAAMAM

Query Match 94.0%; Score 2946; DB 9; Length 3219;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3096; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

BASE COUNT 934 a 696 c 748 g 841 t
 ORIGIN
 EKAAEDTGCIHMEAPETYIEPSRAVSLFFNNKOEFTLEVLTRPFSKLNKODIRLKG
 IESSATSLRLQKRCAGVAGSISLVSTCKNYSIMLWASPLTDEKHTSVNFKT
 LSIHDLONORLPGGLTSDGINKLTKTLMINKNEDAKIKLEGGKMKCLFHL
 THLSIDGGMVYIKVSLSSBPCDDEIDLVSCCLSAANVKTLAONLHVLKSLIDS
 ENYLEKDENEALHELIDRMNVLEDTLALMLPWGCVQGSLSLKHLEEVPLQVKGIL
 KNWRLTDEIRILGAFENRNPKNFOQNLNLRVSSDGMILAFMGVFNELQVLFDFR
 STKEFLPDPAVVRKLSOYLSKTFLEQERLRVGMQFDDDLISVITGAFKLVTYA"

Db 12 ggtgggaagcttccatccgaagaatgaattccataaagaacatagccgaacctatt 71
 QY 12 ggtgggaagcttccatccgaagaatgaattccataaagaacatagccgaacctatt 71
 Db 121 ggTGGAAAGCTTTGATCCGAACAAATGAATTTGATTAAGCAATAGCCGACCTTATT 180
 QY 72 caaagaatggaatgactgtataaagaacaatcacagatgacctatttgatgaatgt 131
 Db 181 CAAAGATGGAAATGACTGTATTAAAGCAATCACAGATGACCTATTGTATGAATGT 240
 QY 132 ctgaatcggaagaagtaaacatcattgtctcgagaaggtggaaggaatgctgctaga 191
 Db 241 CTGAATCGCGAAGAAAGTAACATCATTTGCTCGAAGAGGTGAGAGGATGCTGTAGA 300
 QY 192 gggatcattcacatgattttgaaaaagggttcaagagctctgtaacctcttttaattcc 251
 Db 301 GGGATCATTCACATGATTTTGAANAAGGTTTCAGAGTCTGTAACTCTTTTAAATCC 360
 QY 252 ctlaagagtggaactatccatctattctcagaacttgaatggaacagtcctttcatcag 311
 Db 361 CTTAAGAGATGGAATGACTCTCTATTTCAGACTGAATGAGCAAAAGTCTTTTCATCAT 420
 QY 312 acatcaagaaggaacttggagcaattgtgctcaggaatttaagaagacttgaaccacca 371
 Db 421 ACATCAGAAAGAGACTTGAGACATTTGGCTCAGAGATTAAAGACTTGATACATACCCA 480
 QY 372 tctttctgaacttctatcccttggtgaagaattatgacattatttaacttgaagaagc 431
 Db 481 TCTTTCTGAACCTTTATTCCTCTGTGTAAGATATTGACATATTTTTAACTGAAAAGC 540
 QY 432 accctacagaagccgtctgtggaagaaggaaccaacacatcacgcgttgagcagctg 491
 Db 541 ACCCTCACAGAACCTGTCTGTGAGGAAGACCAACACATCACCGCGTGGAGCAGCTG 600
 QY 492 accctgaatgagctctgtgagagctcttcaagagccctgacatattgaaggaggaatctg 551
 Db 601 ACCCTGAATGAGCTCTGTGAGGCTCTTCAGAGCCCTGTGATATTGAAGGGAATCTGGC 660
 QY 552 aaagcaagtcacatctgtctgagcagcatgtgcatgctctggggcttcggaagtgcaag 611
 Db 661 AAAGCAAGTCACATCTGTGTGACGAAGATGTGCATGTCTGTGGGCTCCGGAAGTGCAAG 720
 QY 612 gctctgaccaagtcacaattgctctctctcctcgctcgaagggccagggtggaact 671
 Db 721 GCTCTGACCAAGTTCAAAATTCCTCTTCCTCCGCTTCAGCAGGAGCCAGAGGTGAGACT 780
 QY 672 ttggaacaccttgatgaactcctctgtatataactctgacacataaggaagaagcattc 731
 Db 781 TTTGAAACCTTGATCACTCCGATTAACCTGGATACGACATCAATGAGAGACGACATTCC 840
 QY 732 atggcagatgctgtaagcttgcggaagaggtctctttcctcttgatgctcaaatgaa 791
 Db 841 ATGGCATGCTCTGTAAGCTGCGGACGAGAGGTTCTTTTCTCTTGATGCTCAATGAA 900
 QY 792 ttaagccccaagatgccagaaatcgaaagccctgataaagaagaaacacacgcttcaag 851
 Db 901 TTCAAGCCCAAGACTGCCCAAGAAATCGAAGCCGTGATTAAGAGAAACACACGCTTCAAG 960
 QY 852 aacatggtcatgtaaccactacatgagtgctcgtgaggaacaatcggaggttggtgac 911

D	b	961	AACATGCTATGCTACCACTACCACTGAGTGCCTGAGGACATATACGGCAGTTGGTGCC	1020
O	y	912	ctgaactcgaagtgaggagatacgaagaagacgcgccaggtccatccgaagatg	971
D	b	1021	ctgactgcgtgaggtggggatgatgacagaaagacgcgccaggtcgtcattccgaagtg	1080
O	y	972	ctgataaagaagctctgcgaaggtctgtctccaaattcagaatccagtgctgaag	1031
D	b	1081	ctgattcaagagagcttgcctgaagcctgttgctccaattcagaattccaggtccttgag	1140
O	y	1032	aattctatgaagaacccctctcttctgtgtcatcactgtgtgcaatccagaatgggtgaagt	1091
D	b	1141	aattctatgaagaacccctctcttctgtgtcatcactgtgtgcaatccagaatgggtgaagt	1200
O	y	1092	gaattccactctcaacacacaaacgcgtgtctcaactctcatgctgtgtgaacag	1151
D	b	1201	gagttccactctcaacacacaaacgcgtgtctcaactctcatgctgtgtgaacag	1260
O	y	1152	aaaaacaaacacacataaagaagtgtgtcgtcgaagtactcaatcggagcctgtgacac	1211
D	b	1261	aaaaacaaacacacataaagaagtgtgtcgtcgaagtactcaatcggagcctgtgacac	1320
O	y	1212	tgtgtgagactgagctctgtgaaggtgtgtctcccaaaagtgtgaattcgaactgcagat	1271
D	b	1321	gtgtgagactgagctctgtgaaggtgtgtctcccaaaagtgtgaattcgaactgcagat	1380
O	y	1272	gtgtcagcgttaatgaagaatgtccctgtcgtcaaacctgggtcctcctgttaatatagact	1331
D	b	1381	gtgtcagcgttaatgaagaatgtccctgtcgtcaaacctgggtcctcctgtgttaatatagact	1440
O	y	1332	caaaagtctcaagccaaagataaattctcttcacaagtcattccaggagtlacacagcaga	1391
D	b	1441	caaaagtctcaagccaaagataaattctcttcacaagtcattccaggagtlacacagcaga	1500
O	y	1392	cgaaagactcaagcagttatctgaacgtctcatgaagcagaagaggttgaccaaaggaatggt	1451
D	b	1501	cgaaagactcaagcagttatctgaacgtctcatgaacgtctcatgaacgtctcatgaacgt	1560
O	y	1452	tacttcgagaagaatgtttccattcggacatctacatccactatagcagcgtctccg	1511
D	b	1561	tacttcgagaagaatgtttccattcggacatctacatccactatagcagcgtctccg	1620
O	y	1512	tacacgtgtgtgtcatctgtgtgaagccacagagcgtgttatagaagcactccgcagcagt	1571
D	b	1621	tacacgtgtgtgtcatctgtgtgaagccacagagcgtgttatagaagcactccgcagcagt	1680
O	y	1572	tatcaaacagcgtcgtcctctcgtgacttccatctgcgcgaagaagcctctctgtgaagcaga	1631
D	b	1681	tatcaaacagcgtcgtcctctcgtgacttccatctgcgcgaagaagcctctctgtgaagcaga	1740
O	y	1632	tcttcgcaagcgtgtgaacacacacactgaagcaagaatcttgaagccatataacatcaat	1691
D	b	1741	tcttcgcaagcgtgtgaacacacacactgaagcaagaatcttgaagccatataacatcaat	1800
O	y	1692	tcccttgtatagatgtgtgtcatctcatatatacgaagaagtlacatcccaatccagcctgagc	1751
D	b	1801	tcccttgtatagatgtgtgtcatctcatatatacgaagaagtlacatcccaatccagcctgagc	1860
O	y	1752	caagaattggaagcctctcttccaaagtataaagcttatatacaactccaggaataccccc	1811
D	b	1861	caagaattggaagcctctcttccaaagtataaagcttatatacaactccaggaataccccc	1920
O	y	1812	gattactatattggaactcttgaacattgtcccaattgtgtcaagtgctcgtgactcat	1871
D	b	1921	gattactatattggaactcttgaacattgtcccaattgtgtcaagtgctcgtgactcat	1980
O	y	1872	aaactggaatttlatgtgggagctatgtgtctcatctgtggaagagcgtcagaagaacagct	1931
D	b	1981	aaactggaatttlatgtgggagctatgtgtctcatctgtggaagagcgtcagaagaacagct	2040
O	y	1932	ggaattccacatggaagagcccccagaacactcatctccagcagagcgtgtatctgttc	1991
D	b	2041	ggaattccacatggaagagcccccagaacactcatctccagcagagcgtgtatctgttc	2100

```
Qy      3072 agtgcattacggctgcttttaactcgttaactgcttaa    3110
          |||||
Db      3181 AGTGTTATTACAGGCGCTTAACTAGTAACGCTTAA    3219
```

LOCUS	AY027787	3396 bp	mRNA	PRI	20-JUL-2001
DEFINITION	Homo sapiens CIANA (CIAN1) mRNA, complete cds.				
ACCESSION	AY027787				
VERSION	AY027787.1	GI:14324112			

ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 3396)
Damiano,J.S., Stehlik,C., Plo,F., Godzik,A. and Reed,J.C.
Clan, a novel human ced-4-like gene
Genomics. 75 (1-3), 77-83 (2001)
21365712
11472070
2 (bases 1 to 3396)
Stehlik,C., Damiano,J.S., Plo,F., Godzik,A. and Reed,J.C.
Direct Submission
Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla,CA 92037, USA

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="p22-p21"
/tissue_type="lung"
1..3396
/gene="CIAN1"
277..3351
DSDS

```

CDS

CDS

BASE COUNT ORIGIN	992 a	737 c	793 g	874 t
----------------------	-------	-------	-------	-------

Query Match	93.0%;	Score 2915;	DB 9;	Length 3396;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 3115; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY 12 gggtgggaaccttccatccagacaatgaaattccataaaggacaaatagccgaagccctatt 71
|||||
Dp 253 ,ggcgggaacctttccatccagacaatgaaattccataaaggacaaatagccgaagccctatt 312

QY	72	caagaatgggaatgactggtataaagaacaatacaagatgacacitatttgatgaattt	131
Db	313	CAAGAATGGGAATGACTGTTATTAAGCAATACAGATGACCTATTGTATGAATGT	372
QY	132	ctgaatcgcgaagaagaaacacatcatcttgctgcgaagaagtgtagcaagatgctctaga	191
Db	373	CTGAATCGCGAAGAAGTAAACATCATTTGTGTCGAGAAGGTGGAGAGATGCTCTAGA	432
QY	192	ggagatcatcaatgatattgaaaaaggtlcaagatccctgtaacctcttctaattcc	251
Db	433	GGGATCATTCACATGATTTGAAAAAGGGTTCCAGATCCTGTACCTCTTTCTTAAATCC	492
QY	252	cttaagaagatgggaactatccctctattccgaagcttgaatggacaagatcttttaacg	311
Db	493	CTTAAGAAGTGGAACTATCCCTCTATTTCAGAGATTTGAATGACAAAGTCTTTTTCATAG	552
QY	312	acatcagaagaagagactctgagacgaatttggcctcaagatcttaagaactgttacataccca	371
Db	553	ACATCAGAAGAAGACTTGGACGATTTGGCTCAGAGATTTAAAGACTGTACATACCCCA	612
QY	372	tctttctgaaacttatacccttggltgaagalatgacataattttaacttgaagaac	431
Db	613	TCTTTCTGAACTTTAAATCCCTTTGGTGAAGATTTGACATTATTTTTTAACTTGAATAAGC	672
QY	432	accttcaaaaactgtgcgttggagaagaagacaaacacatcaacgcgttgaagaagt	491
Db	673	ACCTTCAACGAACCTATCTCTGTGAAGAAGACCAACACATACCGCTGAGAGAGCTG	732
QY	492	aaccttgaatggcctctgcgaagctcttccagagccctgcatcatgaaagaagaaactgac	551
Db	733	ACCCTGAATGGCCTCGTGCAGAGGCTTTCAAGAGCCCTGTATCATTTGAAGGGAAATCTGGC	792
QY	552	aaaggaagtcacactctgtcgtcagaagcatgccaatgctcttgggtctcogaagtgcgaag	611
Db	793	AAAGGCAAGTCACCTCTGCTGCAGAGCCATTTGCCATCTCTGGGGCTCCGGAAGTGCAGAG	852
QY	612	gctctgaacaaatccaattccctctctctctccctccctctgaagaagcccaaggttgactt	671
Db	853	GCTCTGACCAAGTTCAATTCGTTCTTCTCTCCGCTCAGAGAGGCCACAGGGTGGACTT	912
QY	672	tttbaaacctctgtgatcaacactccctgatataccctgcgacacatcaagaagaagacatctc	731
Db	913	TTTBAAACCTCTGTGATCAACCTCGAGATTAACCTGGGACATCAGAGACAGACATTC	972
QY	732	atggcaactgctgcgtgaagactgcgcgaagaaggtcttctctctctgatlgtactaaatgaa	791
Db	973	ATGGCACTGCTGCTGTAAGCTGCGGACAGAGGGTCTTTCTCTTGATGGCTACATGAA	1032
QY	792	tttaagcccccaagactggcccaagaataatcgaagccctgataaagaanaaacacgcgttcaag	851
Db	1033	TTTCAAGCCCCAAGACTGCCCAGAATTCGAAGCCCTGTATTAAGGAATAAACCCGCTTCAAG	1092
QY	852	aacaatgaaactgaaacacataccacttgaatggcctgaagacataacgaagattgtgtcc	911
Db	1093	AACATGTCATCGCACACACTACCACTGAGTGGCTTAGGACATACGGAGAGTTTGTTGCC	1152
QY	912	ctgacatgcgaagltggyggataltgaagaagaagacgcgcacagctctcatccagaagaatg	971
Db	1153	CTGACTGCTGAGGGGGGATATGACAGAAAGACAGCGCCACAGGCTCTATCCGAAAGATG	1212
QY	972	ctgatacaagaagactgcttgaagactgttgttcttccaattcagaaaatccaagatggtcttgag	1031
Db	1213	CTGATCAAGAGCTGTGCTGAAGGCTTTGTCCTCAAAATTCAGAAATTCAGAGGTGCTTGAAG	1272
QY	1032	aactcaatgaagaacccctcctcttgttgcataacttgcgaatccagaatggtygaagt	1091
Db	1273	AATTCATGAGAGACCCCTCTCTTTGTGTGTCATCATTGTGCAATTCAGATGGGGTAAAGT	1333
QY	1092	gaattccactctacacacaaacaaacgcgttccataactcttatalgatctgttgaataag	1151
Db	1333	GAGTTCACATCTCACACACAAACAAACGCTGTTCCATACCTTATGATGATCTGTTGATACAG	1392
QY	1152	aaaaacaacacataaagaatgtgtgcttgcgaagtgcattcatcgagacttgacac	1211

QY 1152 aaaaacaacaacataaaggtgtgtcctgcagtgacttcattcggagcctgaccac 1211

Db 1393 AAAAACAACACAAACATFAAGGTGTGGCTGCAAGTACCTTATCGAGCCGAGCCAC 1452
 Qy 1212 tgttgagacctgctctggagggtgtgtctctccacaagtttgatctgcaactgcagat 1271
 Db 1453 CCGTGAAGACCTGCTCTGGAGGGGTGTGTCTCCACAAAGTTGATTTGCAATGCGAGAT 1512
 Qy 1272 gttcagcgtgaatgaggatgtccctgcgcgaactgggctcctcgttaatttaagct 1331
 Db 1513 GTGTCAAGCTTAATGAGGATGCTCGTGCACACTGGGCTCCTGTTAAATATACAGCT 1572
 Qy 1332 caaaggttcaagccaaagataaattctcttcaaaagttcaccagagttacacagcagga 1391
 Db 1573 CAAAGGTTCAAGCCAAAGATATAATCTTTCAACAACTCATTCAGAGATACACAGCAGGA 1632
 Qy 1392 ggaagactcagcagttatctgacgctcactgagcagaggaagtgaccaggggaatgt 1451
 Db 1633 CCAAGACTCAGAGGTTATTTGACGTCTCATGAGCCAGAGAGGTGACCAAGGGGAATGCT 1692
 Qy 1452 tacttgcaaaatgtgttccacttccgacattacatccactatagcagggctgcgcgg 1511
 Db 1693 TACTTGCAAAATATGTTTCCATTTCGACATTACATCCATTATAGCAGCCTGCTCCGG 1752
 Qy 1512 tacacctgtggttcatctgttggaagccacagggctgttaltgaaagacctgcgcagctg 1571
 Db 1753 TACACCTGTGGGTCACTGTGGAAGCCACAGGGCTGTTATGAAAGCACTCCGACGAGTG 1812
 Qy 1572 tatcaaacgctgctcctctgcgacttccatccgccaagaagcctctctggaagcagaa 1631
 Db 1813 TATCAACACAGGCTCTCTCGACTTTCATCGCCAAAGGGCTCTCTGGAACAGGAA 1872
 Qy 1632 tcttgcaaaagtgtgaaaacacacactgagcagaaattctgaagccataaacatcaat 1691
 Db 1873 TCTTTGCAAAAGTGTAAACACCACTGAGCAAGAAATTTCTGAACCCATAAACATCAAT 1932
 Qy 1692 tcccttgtagagtggtgcatccattatatacaagaagatataccaaatcagccctgagc 1751
 Db 1933 TCCTTTGAAGAGTGGCATCATTTATATCAAGAGATACATCCAAATCAGCCCTGAGC 1992
 Qy 1752 caagaattggaagcttcttcaagaagtaaaagctatatatacaactcaggggaatcccc 1811
 Db 1993 CAAGAATTTGAAGCTTTCTTCAGGTAAACCTTATATATCAACACAGGGAACATCCCC 2052
 Qy 1812 gattactatattgacttcttgaacattgcccattgagcagtgctctggaactcatt 1871
 Db 2053 GATTACTTATTTGACTTCTTTAACAATTTGCCAATTTGTCAGAGTCCCTGGACTTCAT 2112
 Qy 1872 aaactggaactttaaagggagctatggtctcatgggaaaaggtctcagaagaacacaggt 1931
 Db 2113 AAACGTGACTTTATGGGGAGCTATGGCTCATGGGAAAGGCTGCAGAGACACAGGT 2172
 Qy 1932 ggaatccacatggaagagggcccccagaaacctacatccccaagggctatcttctgttc 1991
 Db 2173 GGAATTCACATGGAAGAGGCCCCAGAAACCTACATTCACAGAGGCTGTATCTTTGTTC 2232
 Qy 1992 ttcacactggaagcaggaattcaggaactcgtggaagtacacactccgggaattcagcaagt 2051
 Db 2233 TTCAACTGGAAGCAGGAATTCAGGACTCTGAGAGTACACTCCGGGATTTTCAGCAGAGTG 2292
 Qy 2052 aataagcaagatatcacataltctgggaaaataatactgaactctgcacaaagcctcaggt 2111
 Db 2293 AATAAGCAAGATATCAGATATCTGGGGAATAATATTCAGCTGCGCCAAAGCCTCAGGCTG 2252
 Qy 2112 caataaagaagatgtgcgtgtgtgtgctggaagcctcagttgtgtctcagcaactgttaag 2171
 Db 2353 CAAATTAAGAGATGTGCTGGTGTGGTGAAGACCTCAGTTGTGCTCTCAGCAGCCTTAAG 2412
 Qy 2172 aacattatctcctcatgtgtgaagcaggtccctccacataagaagaatgaagagcactc 2231
 Db 2413 AACATTATTTCTCATGTGTGAAGCCACTCCCTACCATGAAATGAAGAGCACATC 2472
 Qy 2233 acatctgaacaaacctgaaacctgagatctatcatgactcagatcaagcagcgcgcg 2291

Db 2473 ACATCTGTAAACAAACCTGAACACCTTGAGTATTCATGACCTACAGATTCAGGCTCCG 2532
 Qy 2292 ggtgtctgactcagcagcttggglaacttgaagaacctataaagctcataatgataac 2351
 Db 2533 GGTGTCTGACTGACACTTGGGGTAACTGTGAAGAACTTACAAAGCTCATATATGATTAAC 2592
 Qy 2352 ataaagaatgaatgaagaagtgctataaactagctgaagggcctgaanaacctgaagaag 2411
 Db 2593 ATTAAGATGAATGAGAGAGATGCTATTAACCTACTTAAGGCTCGAATAAACCCTGAAGAAG 2652
 Qy 2412 atgtgttatttcaacttgagaccactgtctgacacttgagagggagatgatatgactc 2471
 Db 2653 ATGTTTATTTATTTATTTGACCACTTGTCTGACATTGGAAGGGAATGATATATATGCT 2712
 Qy 2472 aagtcctctcgaatgaacctgtgaccttgaagaatccaatagatgctcctgctgtg 2551
 Db 2713 AAGTCTCTGCAAGTGAACCTGTGACCTTGAAAGAAATTCAAATTAATGATGCTGCTGTG 2772
 Qy 2532 tctgcaaatgcaatgaagaatccctagctcagaatcttcaaatgtgtccaactgagact 2591
 Db 2773 TCTGCAAAATCAGTGAATAATCTAGCTCAGATCTTCAAAATTTGGTCAAACTGAGCAT 2832
 Qy 2592 ctgatttatacagaanaattacacttggaanaagatlgaaaatgaagcctctcactgaactgac 2651
 Db 2833 CTGATTTATCAGAAATTTACCTGGAATAAAGATGGAATGAAGCTCTTCATGAACCTGATC 2892
 Qy 2652 gacagaatgaagtgctcagaacagctcaccgactgactgtgctccttggtggtgtgagctg 2711
 Db 2893 GACAGATGAACGTGTGAGAACAGCTCACCGCACTGATGTGCTGCGCTGGGCTGTGAGAGTG 2952
 Qy 2712 caagcagccttgagcaagcctggttgaaaacatttggaggaggttcccaaacctgcataagct 2771
 Db 2953 CAAGCGAGCTTGACAGACCTGTTGAAACATTTGGAGAGAGTCCCAACAATGCTGAACCTT 3012
 Qy 2772 ggggtgaaaaaactggaagactcacagatcacagagatlgaaatttaagttgcatlttttga 2831
 Db 3013 GGGTTGAAAAACGTGAAGACTCACAGATACAGAGATTGAAATTTAGTGTGATTTTGA 3072
 Qy 2832 aagaacctcttgaanaaacttccagcagcttgaatttggcggggaactcgttggagcagtgat 2891
 Db 3073 AAGAACCCCTGTGAATAAATTCACAGCACTTGAATTTGGCGGGAATCTGTGAGCAGAT 3132
 Qy 2892 ggtgtgcttgccctcactgaaggtgtgatttggagaacttgaagaatlaggttttggactt 2951
 Db 3133 GGTGCTTCCCTTCATGAGGTGATTTTGAAGATCTTAAGCAATTAAGTGTGTTTGTGACTTT 3192
 Qy 2952 agtactaaagaatttctactgtatccagcatlagtgcagaanaacttaagccaagtgtatcc 3011
 Db 3193 AGTACTAAAGAAATTTCTACTGATCCAGCATTTAGTGAAGAACTTATACCAAGTGTATCC 3252
 Qy 3012 aagtttaacttcttgcagaagaagcttaggcttgtgtgtgcaatttgaatgatgatctc 3071
 Db 3253 AAGTTAACTTTTGTGCAAGAGCTAGGCTGTGGCGGCAATTTGATGATGATGATGATC 3312
 Qy 3072 aggttatttcaaggtgctttaaactagtaagcagcttaaaagtgatctcgaagcga 3130
 Db 3313 AGTGTATTACAGGTGCTTTTAACATGATGATGATGATGATGATGATGATGATGATGATGAT 3371

RESULT 4
 AF376061 3581 bp mRNA PRI 15-MAY-2001
 LOCUS Homo sapiens caspase recruitment domain protein 12 mRNA, complete cds.
 DEFINITION AF376061
 ACCESSION AF376061.1 GI:14040074
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 3581)
 AUTHORS Gingras, M., Qiu, J. and Margolin, J. F.

TITLE	Differential expression of the caspase recruitment domain protein 12 (CARD12) during monocytic differentiation
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3581)
AUTHORS	Gingras, M., Olu, J. and Margolin, J.F.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAY-2001) Pediatric/Texas Children's Cancer Center, Baylor College of Medicine, 6621 Fannin St. MC3-3320, Houston, TX 77030, USA
FEATURES	Location/Qualifiers
source	1. 3581
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2p21-22"
	/cell_line="U937"
	/cell_type="peripheral blood-derived CD14 mature monocyte"
	490..3564
CDS	/note="CARD12"
	/codon_start=1
	/product="caspase recruitment domain protein 12"
	/protein_id="AAK53443.1"
	/db_xref="GI:14040075"
	/translation="MNFIKDNRALIQRMGMVYIKQITDDLFVWVNLKRENNITCCF KVEDPAAGITIMLIKSGESCNLFKSLKEMNPLFODLNGQSLFHSBDDLDLA ODLKDLYHPSPFLNRYPLGEDIDILFNLSFTFTEVLMRKQHHVRBQLTLNGLOA LQSPCLIEGSGKSKSTLQRIAMLMGSGKALFKEVFLRSRQGGFETLDC QLDIPGTRKOTFAMLLKLRVFLFLMGKALFKEVFLRSRQGGFETLDC VTTTECLRHIFQFALFAEYGDMDTDSQALIRFVLKLEAGLELLOIOSRCRLN MKTPLFVITCAIOGSESPHSHTQTLFHFYDILLKNNKHGKGVASAPDRISLD CGDALRGVHFHKRDEPDYSSVEDALITDGLCKRTAORFKRYKFPKRSPEY AGRLSLLTSHPEEVTNGMYLQKMSISDITSTYSLLRTGSSVEATRAVMKH LAAVQHGCLGLSLAKRPLWRQESLQSVKNTTEQELIKAININSEVCGHILVQEST SKSLSOFEAFPOKSLYINGSNIPDLFDFEPLPCASALDFIKLDFEGGAASW EKAEDTGIHMEAREPYIPSRVSLFENKQEFRTLEVLDRFSKLKODIRYLGK IFSSATSLRIQKRCAGVAGSLVSTCKNYSLWVAPSLTIDEDEHITSTYLTGK LSIHDLQORLPGLTGLSLGKNTKTLKINDIKNEEDAIKLAGLKNKKMCLFHL THLSDIGRMVYIKSLSSEPDDEIDVSCCLSAVKLLAQLNHLVKLSITDLS ENYLEKDSNEALHELIDRMVLEDTALMLPWGCVQSSLSLHLEHVEPQVLVGL KNMRLTDEIRILGAFEGKNPKNFOQLNLAGNRVSDGWLAFMVGFPENLQVYDFD STKEFLPDPALVRKLSQVLSKLTFLQEARLVGWOPDDDLISVITGAFKLVYA"
misc_feature	502..741
	/note="Region: caspase recruitment domain"
ASE COUNT	1033 a 781 c 843 g 924 t
RIGIN	
Query Match	92.9% Score 2912; DB 9; Length 3581;
Best Local Similarity	99.9% Pred. No. 0;
Matches 3112:	Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	12 ggtgggaagcttcattccagaacaatgaattcataaagaacatagccgaagcccttatt 71
DB	466 ggtgggaagcttcattccagaacaatgaattcataaagaacatagccgaagcccttatt 525
QY	72 caaagaatgggaatgactctgtataaagaacaatcagaatgacctattgtataggaaatgtc 131
DB	526 caaagaatgggaatgactctgtataaagaacaatcagaatgacctattgtataggaaatgtc 585
QY	132 ctgaatcgcgaagaatgaacatcatcttgctcgagaagaatggagaagaatctgctgata 191
DB	586 ctgaatcgcgaagaatgaacatcatcttgctcgagaagaatggagaagaatctgctgata 645
QY	192 gggatcatcacatgatttgaaaaaagggttcagaagtcctgttaacctcttcttaaatcc 251
DB	646 gggatcatcacatgatttgaaaaaagggttcagaagtcctgttaacctcttcttaaatcc 705
QY	232 cttaaggagtggaactatccctctatttcagaactgaaatgagcaaaagtccttttcaatcag 311
DB	706 cttaaggagtggaactatccctctatttcagaactgaaatgagcaaaagtccttttcaatcag 765
QY	312 acataagaaggagacttgagacattggtcaggaatttaaggagactgtacatacccca 371
DB	766 acataagaaggagacttgagacattggtcaggaatttaaggagactgtacatacccca 825
QY	372 tcttctcgaactttatccctctgtggaagatattgacataatttttaacttgaaagc 431
DB	826 tcttctcgaactttatccctctgtggaagatattgacataatttttaacttgaaagc 885
QY	432 accctcaagaacccgtctgtggaaggaagcaaacacatcacccgtgtgagacgtc 491
DB	886 accctcaagaacccgtctgtggaaggaagcaaacacatcacccgtgtgagacgtc 945
QY	492 accctgaatgagctctctggaagctcttcaagaccccttgatcatatgaagggaatctgac 551
DB	946 accctgaatgagctctctggaagctcttcaagaccccttgatcatatgaagggaatctgac 1005
QY	552 aaaggcaatccactctgtcagcgatctgcatgctctgtgaggtccggaagtgcag 611
DB	1006 aaaggcaatccactctgtcagcgatctgcatgctctgtgaggtccggaagtgcag 1065
QY	612 gctctgacaaatgtaaatgcttcttctcgtcgtlctcagaagggccaggtgagact 671
DB	1066 gctctgacaaatgtaaatgcttcttctcgtcgtlctcagaagggccaggtgagact 1125
QY	672 ttggaaccccttgatcaactcctctgataataactgacacaataagaagaagacatc 731
DB	1126 ttggaaccccttgatcaactcctctgataataactgacacaataagaagaagacatc 1185
QY	732 atgagcatgctgtcgaagtcgagcagaaggtctcttctcctctgatagtctacaaatga 791
DB	1186 atgagcatgctgtcgaagtcgagcagaaggtctcttctcctctgatagtctacaaatga 1245
QY	792 ttcagaagcccaagactgtcccaagaatcgaaagccctgataaaggaataaccacgcgtc 851
DB	1246 ttcagaagcccaagactgtcccaagaatcgaaagccctgataaaggaataaccacgcgtc 1305
QY	852 aacatggtcatgctgtaac 911
DB	1306 aacatggtcatgctgtaac 1365
QY	912 ctgactgctgaagtgaggatataagcaagaacagcgccaggtctcattccgaagatg 971
DB	1366 ctgactgctgaagtgaggatataagcaagaacagcgccaggtctcattccgaagatg 1425
QY	972 ctgatacagaagactgtgtaaggtgtgtgtcccaaatccagaatccaggtgtgag 1031
DB	1426 ctgatacagaagactgtgtaaggtgtgtgtcccaaatccagaatccaggtgtgag 1485
QY	1032 aatcctatgaagaccctctcttctgtgcatcaactgtgcaatccagaatggtgaaagt 1091
DB	1486 aatcctatgaagaccctctcttctgtgcatcaactgtgcaatccagaatggtgaaagt 1545
QY	1092 gagttccactctcac 1151
DB	1546 gagttccactctcac 1605
QY	1152 aaaaac 1211
DB	1606 aaaaac 1665
QY	1212 tgtgagacctagcctctgaggtgtgtgtctccacaagtttgatcttgaacttgagat 1271
DB	1666 tgtgagacctagcctctgaggtgtgtgtctccacaagtttgatcttgaacttgagat 1725
QY	1272 gttccagcgatgaatgaagatgttccctgtgcaaacatgggtctcctgtlaatalatacag 1331
DB	1726 gttccagcgatgaatgaagatgttccctgtgcaaacatgggtctcctgtlaatalatacag 1785
QY	1332 caaaggttcaagccaaagtataaatcttctcaagatcatccagaggtgatacagaagga 1391
DB	1786 caaaggttcaagccaaagtataaatcttctcaagatcatccagaggtgatacagaagga 1845
QY	1392 cgaagactcagcgtttatctgacgtctcctcctcagcaagaagaggttgaacaaagggaatgt 1451
DB	1846 cgaagactcagcgtttatctgacgtctcctcctcagcaagaagaggttgaacaaagggaatgt 1905

```

OY 1452 tacttcagaaaaatggtttccatttcggaacattacatccactatagcagcctgtccg 1511
    |||
Db 1906 TACTTCGCAAAATGTTTCCATTTCGGACATTACATTCATTATAGCAGCTGCTCCG 1965
OY 1512 tacacgtgtggtcaltctgtgaagcaccagggcgttatagaagcactcgaagctg 1571
    |||
Db 1966 TACACTGTGGGTCATCTGTGGAAAGCCACAGGCGTGTATGAAGACCTCGAGCAGTG 2025
OY 1572 tatcaacagcgtcctcctcgtgaacttccatcgccaagaagcctctctggaagaa 1631
    |||
Db 2026 TATCAACAGCGCTGCTTCGACCTTCATCGCCACAGAGGCTCTCTGAGACAGGAA 2085
OY 1632 tcttcgcaagtggtgaaacaccacactgagcaagaattctgaagccataaacat 1691
    |||
Db 2086 TCTTTGCAAAAGTGTGAAAAACACACTGACAGCAAAATTCGAAAGCCATTAACATCAAT 2145
OY 1692 tcccttctagagtggtgcaatccatatacaagaagatccatccaaatcagcctgagc 1751
    |||
Db 2146 TCCCTTGTAGAGTGTGGCATTCATTATATCAAGAGATACATCCAAATCAGCCCTGAGC 2205
OY 1752 caagaatttgaagcttcttcaaggtaaaagctatatacaactcagggaaatcccc 1811
    |||
Db 2206 CAAGAAATTTGAAGCTTCTTCAAGGTAAAGCTTATATATCAACTCAGGGAACATCCCC 2265
OY 1812 gattacttaattgactctcttgaacatttgcacaaattgtgcaagtgcttgaactcatt 1871
    |||
Db 2266 GATTACTTATTTGACTCTCTTTTAAACATTTGCCCAATTTGCAAGTGCCTGGACTTCATT 2325
OY 1872 aaactggaacttataaggggaagctatgtctcatgtggaagaaagcgtcgaagaacacagct 1931
    |||
Db 2326 AAACGTGACATTTTATGGGGGAGCTATAGGCTTCATGGGAAAGGCTCAGAAAGCACAGGT 2385
OY 1932 gaaatcccaatgtgaagagagcccccagaacactatcccaagagggctatcttctgttc 1991
    |||
Db 2386 GGAATTCACATGTGAAGAGCCGCCAGAAACCTACATTCACAGAGGCTGTATCTTTGTTC 2445
OY 1992 ttcaactggaagcagaatctcaggaactctggaagtcacactccgggaatttcagaagctg 2051
    |||
Db 2446 TTCACATGGAACAGGAATTCAGAGACTCTGAGAGTACACTCGGAGATTTTCAGCAAGTTGG 2505
OY 2052 aataagcaagatatacatatcttggggaataatatacagctctgccaagcctcagcgtg 2111
    |||
Db 2506 AATTAAGCAAGATATCAGATATCTGGGGAATAATATTACGCTGCGCCAGAGCTCAGGCTG 2565
OY 2112 caaataaagaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2171
    |||
Db 2566 CAATAATGAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2625
OY 2172 aacatttatctctcatgtgtgaagcagctccctcacatagaagaatagagggcacatc 2231
    |||
Db 2626 AACATTATATCTCTCATGTGTGGAAGCCACTCCCTCATCATTAGAGATGAGAGGCATATC 2685
OY 2232 acatcgtatacaaaccttgaaaccttgaatlatcatgacctacagaatcaagcgtgtccg 2291
    |||
Db 2686 ACATCTGTATCAAACTGTGAAGACCTTGATATTCATGACCTACAGAAATACAGGCTGCCG 2745
OY 2292 ggtgtgtctacgtcagcagcctgtgtgaacttgaagaacttacaagaagtctaatgttaac 2351
    |||
Db 2746 GGTGTGTCTACTGACAGCTTGGGTAACTTGAAGAACCTTACAAAGCTCTAATAGATTAAC 2805
OY 2352 atcaagaatgaagaagaatgctataaactagctgaagcgtcgaagaactgaagaag 2411
    |||
Db 2806 ATTAAGATGAATGAGAGAAATGCTATTAATAACTAGCTGAAGGCTGAAAAAAGCTGGAAGAG 2865
OY 2412 atgtgttattatcattgaccactgtctgcaattggaaggaatgattacatagtc 2471
    |||
Db 2866 ATGTGTATTTATTCATTGTGACCACTTGTCTGACATTTGAGAGGGAATGATTAACATAGTC 2925
OY 2472 aagttctgtcgaaggaagcctgtgaaccttgaagaatcaattgctcctgctgtgtg 2531
    |||
Db 2926 AAGTCTCTGTCAAGTGAACTGCTGACCTTGAAGAAATTCATTAAGTCTCTGCTGCTGG 2985

```

```

OY 2532 tctgcaaatgcagttgaanaatcctagctcagaactctcacaaatttggtcacaactgagcatt 2591
    |||
Db 2986 TCTGCAAAATCACTGAAATTCCTTAGCTCAGAAATCTTCACATTTGGTCAAACTGAGCATTT 3045
OY 2592 ctgattatlatcagaanaattacacttgaanaaagaatggaanaatgagcctctcaagctgac 2651
    |||
Db 3046 CTGTGATTTATCAGAAATATTCCTGGAAGAAAGATGGAATGAACCTCTTCATGAACGATC 3105
OY 2652 gacaagatgaagtgcttagaagaagctcaccgcaactgactgtccctgtggagctgtgacgtg 2711
    |||
Db 3106 GACAGATGAACCTGCTAGAACAGCTCACCGCACTGATGTGCTGCGGCTGTGACGCTG 3165
OY 2712 caagcagcctgtgacagcctgttgaagaacttgaagaagcttcccaaacctcgcgaagctt 2771
    |||
Db 3166 CAAGGACGCTTGACAGCCTGTTGAAACATTTGGAGAGGCTCCACACACTGTCAGACTT 3225
OY 2772 ggggtgaaanaactggaagactcacagatatacagaagattgaatttgaagtgacatttttga 2831
    |||
Db 3226 GGCTTGAANAAGTGGAGACTCACAGATACAGAGATTAGAAATTTAGGTGATTTTGGGA 3285
OY 2832 aagaacccctcgaanaaactcagcagcttgaatttggcgggaatcgtgtgagcagtgat 2891
    |||
Db 3286 AAGAACCTCTGAAAAAAGCTTCCAGAGTTGAATTTGGCGGAATGCTGTGACAGATGAT 3345
OY 2892 ggaatgctgacctcactgagtgatgtatgtgaagaactcctaagaatlagtgttttgaacttt 2951
    |||
Db 3346 GGATGGCTTGCTTCATGGGTGTATTTGAGAAATCTTAAGCAATTTAGTGTTTTGACTTT 3405
OY 2952 agtactaaagaatttctacactgataccagcatltagtcaagaanaacttgaagcagtgatcc 3011
    |||
Db 3406 AGTACTAAAGAAATTTCTACTGATCCAGCATTTAGTGTGAAGAACTTAGGCCAAGTGTATTC 3465
OY 3012 aagttacttcttgcagaagaagctagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3071
    |||
Db 3466 AAGTTAATCTTTCTGTGAAGAGAGCTAGGCTGTGTGGGTGAATTTGATGTGATGATGATC 3525
OY 3072 agtgtattacaagtgctttaaactagtaactgactgaacttaataaagtactgaag 3127
    |||
Db 3526 AGTGTATTACAGGTGCTTTTAACATGATGATGATGATGATGATGATGATGATGATGATGAT 3581

```

```

RESULT 5
AC010968
LOCUS AC010968 160583 bp DNA HTG 18-AUG-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-9302, WORKING DRAFT SEQUENCE,
11 unordered pieces.
ACCESSION AC010968
VERSION AC010968.5 GI:9845170
KEYWORDS HTG; HTGS_PHADEL; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160583)
REFERENCE 1
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
2 (bases 1 to 160583)
AUTHORS Waterston, R.H.
REFERENCE Direct Submission
2746 GGTGTGTCTACTGACAGCTTGGGTAACTTGAAGAACCTTACAAAGCTCTAATAGATTAAC
JOURNAL Submitted (28-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108 USA
On Aug 18, 2000 this sequence version replaced gi:8439959.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0093002
----- Summary Statistics -----
Sequencing vector: M13; 59%

```

Sequencing vector: plasmid: 41%
 Chemistry: Dye-Primer ET; 47% of reads
 Chemistry: Dye-terminator Big Dye; 53% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: bases at least Q40
 Consensus quality: bases at least Q30
 Insert size: 147000; agarose-fp
 Insert size: 159583; sum-of-contrigs
 Quality coverage: 6.64 in Q20 bases; agarose-fp
 Quality coverage: 6.38 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently consists of 11 contrigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contrigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1080: contrig of 1080 bp in length
 * 1081 1180: gap of unknown length
 * 1181 2476: contrig of 1296 bp in length
 * 2477 2576: gap of unknown length
 * 2577 5035: contrig of 2459 bp in length
 * 5036 5135: gap of unknown length
 * 5136 9606: contrig of 4471 bp in length
 * 9607 9707: gap of unknown length
 * 9707 17078: contrig of 7372 bp in length
 * 17079 17178: gap of unknown length
 * 17179 27158: contrig of 9880 bp in length
 * 27159 27259: gap of unknown length
 * 27259 45137: contrig of 17879 bp in length
 * 45138 45237: gap of unknown length
 * 45238 65522: contrig of 20285 bp in length
 * 65523 91498: gap of unknown length
 * 91499 91598: contrig of 25876 bp in length
 * 91599 116835: gap of unknown length
 * 116836 116936: contrig of 25237 bp in length
 * 116936 160583: gap of unknown length
 * 160583 43648 bp in length.

FEATURES

SOURCE

1. 160583
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone="RP11-9302"
 1. 1080
 /note="assembly_name:Contig19"
 1181. 2476
 /note="assembly_name:Contig12"
 2577. 5035
 /note="assembly_name:Contig13"
 5136. 9606
 /note="assembly_name:Contig14"
 9707. 17078
 /note="assembly_name:Contig15"
 17179. 27158
 /note="assembly_name:Contig16"
 27259. 45137
 /note="assembly_name:Contig17"
 45238. 65522
 /note="assembly_name:Contig18
 clone_end:17
 vector_side:right"
 65623. 91498
 /note="assembly_name:Contig19"
 91599. 116835
 /note="assembly_name:Contig20
 clone_end:SP6
 vector_side:right"
 116936. 160583
 /note="assembly_name:Contig21"

BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others
 ORIGIN

Query Match 62.1%; Score 1946; DB 2; Length 160583;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1996; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 298 gctcttccatcagacatcagaagaagacttggagacttggcctcgaagattaaagact 357
 Db 49188 GCTCTTTCATCAGACATCAGAGAGACTTGGACATTTGGCTCAGATTTAAAGACT 49247
 OY 358 tttaccatacccccattcttctgaactttalccccttggtygaagatatgtacatttc 417
 Db 49248 TGTACCATACCCCATCTTTTGTGAAGCTTTATCCCTTGTAAGATATTGACATTATT 49307
 OY 418 ttaacttgaagaagcccttcagaagaacctgtctgttgaagaagaacaaacccatccc 477
 Db 49308 TTAAGCTTGAAGACACCTTCACAGAACCTGTCTTGTGAAGAGAACCAACCACTACAC 49367
 OY 478 gcttgaagcagctgaagcccttgaatgacctcctcgaagctcttcagaagccctgaatcat 537
 Db 49368 GCGTGAGCAGCTGACCTGAAATGGCCTCTCTCAGAGCTCTTACAGAGCCCTGCATCATTTG 49427
 OY 538 aaggggaatctgcaagaagcagctcagctcagctcagctcagctcagctcagctcagct 597
 Db 49428 AAGGGGAATCTGGCAAGGCAAGTCCACTGCTCAGCAGCTTCATGCTTGGGCT 49487
 OY 598 ccggaagaatgaaagctctgagcccaagctcaaatctgtcttctcctcgtctcgaaggg 657
 Db 49488 CCGGAAGTGCAGAGGCTCTGACCAAGTTCAATTCCTCTCTCTCCGCTCTCACAGAGG 49547
 OY 658 ccgaaggtggaacttttgaaccctctgtgatacactcctgatatctgcaacaatca 717
 Db 49548 CCCAGGGGAGACTTTTGAANCCCTGTGATCAACTCCTGGATTTACTGGCAATATCA 49607
 OY 718 ggaagcagacatcactgagccatgctgtcgtgaagctgagcagaagggcttcttctctctg 777
 Db 49608 GGAACGACACATTCATGCGCATGCTGTAAGCTCGGACAGGGCTCTCTTCTCTCTCT 49667
 OY 778 atggtacatagaatgaatgaagcccaagactcgcgaagaatcgaagccctgtataagggaa 837
 Db 49668 ATGGCTACAAATGAATTCAGGCCCAAGACTGCCCCGAATTCACACCCCTGATAAGGAAA 49727
 OY 838 accacgcgttcaagaacatggtcaltcgtcaccactaccactcagctgagcagacatc 897
 Db 49728 ACCACGCTTCAAGAACATNGTCATGTCACACATGACACATGATGAGTCCAGGACATAC 49787
 OY 898 ggcagcttgggtgcccctgactgctgaggtggggatagacagaagaagcgccagctc 957
 Db 49788 GGCAGTTTGGTGCCTGACTGCTGAGTGGGGATATGACAGAAAGAGCCGCCAGGCTC 49847
 OY 958 tcatccgaagaatgctgtgtaagaagactgctgtaagagctgtgtgttgcctcaaatcgaat 1017
 Db 49848 TCATCCGAAGATGCTGATCAAGAGCTTGTGAAGGCTTGTGTGCTCCAAATTCAGAAAT 49907
 OY 1018 ccaagttgctgaggaatcactgaagaacccctcctctgtgtgtaactcactgtggaatcc 1077
 Db 49908 CCAAGTGTGAGGAATTCATGAGAACCCCTCTTGTGTGATCACTTGTGCAATCC 49967
 OY 1078 agatggtgtaagatgagttccactcctcaacaacaagaagcgtgttccataactctatg 1137
 Db 49968 AGATGGGTGAAGATGAGTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 50027
 OY 1138 atctgttgaatcagaagaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 1197
 Db 50028 ATCTGTGATPAGAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 50087
 OY 1198 ggaagcctggaacactgctggaagcctcagctcaggaaggtgtgttccccaagtttgaat 1257
 Db 50088 GGAAGCTGAGACACATGTTGAGAGACTAGCTGTGAGAGGCTGTGTCTCCCAAGATTGATT 50147
 OY 1258 tcgaactcaggaatgtgtccagcgltgaatgaggaatgtcctcgtgacaactggtcctct 1317

Tue Mar 26 11:35:48 2002

us-09-697-089-1.oligo.rge

|||||
Db 50148 TCGAAGTCTGAGATGTGCTCCAGCTGATAGATGATGCTGCTGACAACTGGGCTCCTC 50207
OY 1318 gtaataatcacagctcaaaagttcaagccaaafataaatcttcccaagttcaatccagg 1377
Db 50208 GTAAATATACACTCAAGGTTCAAGCCAAAGTAAATTTCTTTCACAGTATTCAGG 50267
OY 1378 agtaacacagagcaagaactcagcagttatgaagcttcctgaagccagagaggtga 1437
Db 50268 AGTACACAGACAGACAAACCTCAGCAGTTATGACCTCATGAGCCAGAGAGAGTGA 50327
OY 1438 ccaagaggaatgttacttcgagaaatgggttcattccgagcattacatcatatata 1497
Db 50328 CCAAGGGAATGGTTACTTGCAGAAAATGGTTCCATTTCAGCAATTAATCATCTTA 50387
OY 1498 gcaagcctgctccggtacacactgtggtgcatctgtggaagccacagaggtgtatagc 1557
Db 50388 GAGGCTCTCCGGTACACTGTGGGTCACTGTGAGAACCCAGGGCTGTATGAGAC 50447
OY 1558 accctgcagcagtgatcaacaagctgctcttcggaattccatcgccagagagcctc 1617
Db 50448 ACTGCGACAGCTGTATACACAGGCTGCTTCGACCTTCATCGCCAGAGAGGCTC 50507
OY 1618 tctgagagaggaatcttgcgaagtgtggaaaacacacactgagcaagaatctgaag 1677
Db 50508 TCTGAGACAGGAATCTTGGAAAGTGTGAAAACACACACTGAGCAAGAAATCTGAAG 50567
OY 1678 ccataaacatcaatctctctgtagagtggtgacatcttatacaagagagatca 1737
Db 50568 CCATAAACATCAATCTCTTGTGAGAGTGGCATCATTTATTCAGAGAGTACATCA 50627
OY 1738 aatcagccctgagccaaagattgaagcttcttcaaggtataaagcttatataact 1797
Db 50628 AATCAGCCCTGAGCCAGAAATTTGAAGCTTTCTTCAAGTAAAGCTTATATATCACT 50687
OY 1798 caggaagaacatcccgatcttacttacttgaactcttgaacatttgcacattgtgca 1857
Db 50688 CAGGGAACATCCCGATTAATTTGACCTTTTGAACCTTGGCCAAATTTGGAAGTG 50747
OY 1858 ctctggaactcaatcaatcagcacttcttgaagagagctatggtctcataggaagctg 1917
Db 50748 CTCTGACCTTCATTAACAGCGACTTTATGGGAGAGCTATGGCTTCAATGGGAAAGCTG 50807
OY 1918 cagaagaacacaggtgagatcccatggaagagagcccgaaacactacattccagagag 1977
Db 50808 CAGAAAGACAGGTGGAATCCATGAGAGAGGCCCCAGAAACCTACATTCACAGAGG 50867
OY 1978 ctgtatcttcttctcaactggaagcaggaattcaggaactctgaggttcaactccgg 2037
Db 50868 CTGTATCTTCTTCACTGAGAGAGAGAAATTCAGGACTGTGAGGCTCACACTCCGG 50927
OY 2038 attcagcaagttgaataagaagatacatcatctgaggaatgaatattcaactctgca 2097
Db 50928 ATTTCAGCAAGTTGATTAAGCAAGATATCTGAGGGAATAATTTACACTCTGCGCA 50987
OY 2098 caagcctcagcctgcaaatgaagagatgctgtgtgtgtgtgtgtgtgtgtgtgtgt 51047
Db 50988 CAAAGCTCAAGCTGCAAAATGAAGAGATGTCTGTGTGTGTGTGTGTGTGTGTGTGT 51107
OY 2158 tcaagcctgtaagaacattatctctcatgtgtggaagccagtcctccatcatagaag 2217
Db 51048 TCAGCCCTGTAAAGCAATTTATCTCTCATGTGTGAGGAGCCAGTCCCTCCATGAG 51107
OY 2218 atgagagacacatcatctgtgaacaacctgaaacacttgatattcatgactacaga 2277
Db 51108 ATGAGAGGCACATCATCTGTGAACAACCTGAAACCTTGTGTGTGTGTGTGTGTGTGT 51167
OY 2278 atcaagagctccgggt 2294
Db 51168 ATCAAGGCTGCGGGT 51184

CNS01D53/c
LOCUS
DEFINITION
BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 164M19 of C19B.978.SKB library from chromosome 2 of Homo sapiens (Human), complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Human
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS
Hazen, J., Fomnechten, N., Mavel, D., Paternotte, C., Sanson, D., Attienave, F., Davoine, C. S., Craud, C., Durr, A., Wincker, P., Brothier, P., Catolico, B., Barbe, V., Burgunder, J. M., Prud'Homme, J. F., Bricle, A., Fontaine, B., Hellig, R. and Weissenbach, J.

TITLE
Spastin, a novel AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Mar 6, 2000 this sequence version replaced g1:602386.

FEATURES
source
1..138909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone.lib="C19B-978-SKB"
/clone.lib="164M19"

BASE COUNT
ORIGIN
39243 a 28424 c 29121 g 42121 t

Query Match
Best Local Similarity 58.9%; Score 1844; DB 9; Length 138909;
Matches 1994; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 298 gttctttcatagacatcagaagagagacttgagagattggtctgagattgaagact 357
Db 75309 GTCTTTTCATCGACATCAGAGAGAGACTTGGACGATTTGGCTCAGATTTAAAGACT 75250

OY 358 tgtacatcacatctcttctgagacttctacccctgtgtgaagatatgtgactatatt 417
Db 75249 TGTACATACCCCATCTTTCTGACTTTTANCCCTGTGGAAGATATGACATTAATT 75190

OY 418 ttaacttgaagaacacattacagaacactgtctgtgtggaagagacacacacatcacc 477
Db 75189 TTAACCTTAAAGACACTTCACAGAACCTGTCTGTGAGAGAGAGACCAACCATCACC 75130

OY 478 gctgtgagacagctgacacctgaatgacctctgagagcttcaagagccctgacatctg 537
Db 75129 CGGTGAGAGAGTGAAGCTTGAATGAGCTCTGAGAGCTCTTCAGACCCCTGACATATTG 75070

OY 538 aaggggaatctgcaaaagcaagtcacactgtctgtgagcagcattgcatgtcttgaggtc 597
Db 75069 AAGGGGAATCTGCAAAAGCAAGTCACTGCTGCGAGCAAGATTCGATGCTGTGGGCT 75010

OY 598 ccggaagtgcaagagctgacaaagtcaaatctgtcttcttcttcttcttcttcttcttct 657
Db 75009 CCGGAAGTCAAGAGCTGTGACCAAGTTCAAAATGTGCTTCTTCTCTCTCTCTCTCTCT 74950

OY 658 ccagaggtgacttttgaagacacctgtgatcaactctcgagatatccctggagacaatca 717
Db 74949 CCCAGGCTGACCTTTTGAACCTCTGTGATCACTCTCTGATATACCTGTGCAACATCA 74890

OY 718 ggaagacacatcatgagcagctgtgaaactctgagagaggttcttcttcttcttcttct 777
Db 74889 GGAACGACATTCATGAGCATGCTGTGAAGCTGGGAGAGGAGTCTTCTTCTCTCTCT 74830

/db_xref="taxon:9606"
/chromosome="2"
/clone="IMAGE CDNA clone 2005417"
/clone_1ib="NCI-CGAP_Pan1"
/map="2p21-p22"
<1..1321
/codon_start=2
/product="hypothetical protein, weakly similar to (AF102871) neuronal apoptosis inhibitory protein 2 [Mus musculus]"
/protein_id="CAB97523.1"
/db_xref="GI:9367840"
/translation="INSGNIPDLYFDFEHLPCNASALDFIKLDFYGAMASWEKAE DRGGIHEAPEPTIIPSRALVLPFNKQEPRLTEVLRPFSLNKQDIRYLGIESSA TSLRIQKRCACVAGSLSTVSTCKNITYSLMPEASLTIEDERHITSVNLKLSIHD LONORLPGLTDSLGKMLKTLKINDINIMNEBDAIKLEGLNKKMCLFHLTILSD IEGMDYIVKSLSSPECDLEEIQLVSCCLSANAVKTLAONLHNLVSLIDLSSENLE KGNENALHELIDRMNVLLEQLTALMLPMGCDVQGSLSLKHDEYVQLKGLKNRL PDTEIRILGAFGKNPLKNFQDLNLAGNVSISDGMVAFMGVEFNKQLVFEFDFSTKEF LPDPALVRKLSQVLSKLTFLQAFARLVGMQFDDDDLSVITGARKLVTAA"

BASE COUNT 416 a 265 c 312 g 362 t
ORIGIN

Query Match 41.2%; Score 1290; DB 9; Length 1355;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1340; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1790 tatcaactcaggagaacatcccgattactattgactcttctgaacattgccaattg 1849
1 TATCAACTCAGGAGACATCCCGATTACTATTGACTCTTGAACATTGGCCCAATTG 60
DB 1850 tgcgaatgctctgacttatttaactgactcttattggaggagctatgctcattgga 1909
61 TGCGAATGCTCTGACTTACTTAACTGACTTTTATGGGGAGCTATGGCTTCATGGGA 120
DB 1910 aaagctctgcagaagacacaggtgaaatccacatggaagagcccgaaacacattacc 1969
121 AAAGCTCGCAGAGACACAGGTGAATCCACATGGAAAGAGCCCGCAAACTCAANTCC 180
QY 1970 cagcagggctgtatcttcttcttcaactggaagcaggaattcaagaactctggaagtcac 2029
181 CAGCAGGGCTGTATCTTCTTCTTCAACTGGAAGCAGGAATTCAGGACTCTGGAGGCTAC 240
QY 2030 accccggagtttcaagaagttaagaagaatlatcacatctctgagggaatatctac 2089
241 ACTCCGGGATTTTCAGCAAGTTGAATTAAGCAAGATATCAGATATCTGGGAAAAATATTCA 300
QY 2090 ctctgcacaagcctcagcgtgcaataaagaatgtgtctgtgtgctggaagcctcag 2149
301 CTCTGCACAAAGCCTCAGGCTGCAATTAAGATATGCTGTGTGCTGGAGCCTCAG 360
QY 2150 ttgtgtctcagcactgttaagaacattatctctcaatggtggaagccagtcacctac 2209
361 TTGTGTCTCAGCAGCCTGTGAAGCACTTATCTCATGTGTGAAGCCAGTCCCTCCAC 420
QY 2210 catagaagaatgagaagcacatcatctgtacaacaaacctgaacacctgtgtattatga 2269
421 CATAGAAGATGAGAGCACATCTGTAAACCTGAACCTGTGAATTTATGA 480
QY 2270 cctacagaatcaacagcgtccgggtgtctgactgacagcttggtgaacttgaagaacct 2329
481 CCTACAGAATCAACAGGCTGCCGGGTGTCTGACTGACAGTTGGTAACCTTGAAGAACT 540
QY 2330 tacaaagtcatatgataaataaagaatgaaatgaagaatgtcatalaactagctga 2389
541 TACAAAGCTCATATGATTAACATTAAGATGAAGATGAAGATGATTAATAACTACTACGA 600
QY 2390 agggccgaaanaaacctgaagaagaatggtttatttcaattgaccacttgcctgaacttga 2449
601 AGGCCTGAAAACCTGAAGAAGATGTTTATTTCATTGACCACTGCTGCTACATGG 660
QY 2450 agagggaaatgaltatagtaagtcctctgtaagtgaaacctgtgaccttgaagaat 2509

|||||
DB 661 AGAGGGAATGATTAATTAATCAAGTCTGTCAAGTGAACCTGTGACCTTGAAGAAT 720
QY 2510 tcaattagtcctctgctgtctgtctcgaatgcaagtgaaatccctagctcaagaattca 2569
DB 721 TCAATTAGTCTCCGTGCTGTCTGTGCAAAATGAGTGAATAATCTTACATGATCTTCA 780
QY 2570 caattggtcaaacctgagcattctctgtattatcaagaataactctggaanaaatgtaaa 2629
781 CAATTGCTCAACTGAGCAATCTTGTATTAACAAGAAATTAATCTGAAAAAGATGGA 840
QY 2630 tgaagctcttcataagacatgacgaagatgaaacgtgtcagaacagctccacgacatgat 2689
841 TGAAGCTCTTCATGAATGATGACAGATGAGACAGTCTTGAACAGCTCAACCGACATGAT 900
QY 2690 gctgcacctgggctgtgacgtgcaagcagcctgagcagcctgttgaacatttgaagga 2749
901 GCTGCCCTGGGGCTGTGACCTGCAAGCAGCAGCTGACAGCCTGTGAAACATTTGGAGGA 960
QY 2750 ggtcccaacactcgtcaagctgtgtgtaaaacttgaagactacagatcagaatgag 2809
961 GGTCCCAACACTGTCAGGCTTGCGTGAATAAAGTGAAGTCAACAGATACAGAGATTAG 1020
QY 2810 aattttagtgcatatttttgaagaacaccttgaanaacttccagcagcttgaattgac 2869
1021 AATTTTAGTGTCATTTTGTGGAAGAACCCTGTGAATAACTTCCAGCAGCTTGAATTTGGC 1080
QY 2870 gggaaatcgtgtgacgaagatgaaatgactgtccctcattgagtgatttgaagaattca 2929
1081 GGGAAATCGTGTGACGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 2930 gcaatttagtcttttgaacttgaacttgaagaatctcctcgaatccagcattgacag 2989
1141 GCAATTAGTCTTTTGTGACTTGTACTTAAGAAATTTCTACCTGATCCACATTAGTCAG 1200
QY 2990 aaactttagcagaagtgtaaccagatgaacttctcgaagaagcctgagctgttgggtg 3049
1201 AAACCTTAGCCAGAGTGTATCCAAAGTTAACTTTCTGCAAGAAAGCTGATGGCTG 1260
QY 3050 gcaattttagtgaatgactcagtgatctatctacagtgctttaaactagtaactgctta 3109
1261 GCAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 3110 aatlaagtgactcgaagcca 3130
DB 1321 AATTAAGTGTACTCGAAGCCA 1341

RESULT 8
AY027788
LOCUS AY027788 1395 bp mRNA PRI 20-JUL-2001
DEFINITION Homo sapiens CLANB (CLANL) mRNA, complete cds.
ACCESSION AY027788
VERSION AY027788.1 GI:14324114
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1395)
AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE Clon, a novel human ced-4-like gene
JOURNAL Genomics. 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
REFERENCE 2 (bases 1 to 1395)
AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source 1..1395
location/Qualifiers


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="Lung"
1..1395
/gene="CLAN1"
277..1356
/gene="CLAN1"
/codon_start=1
/product="CLANB"
/protein_id="AAK14777.1"
/db_xref="GI:14324115"
/translation="MNFIKDNRSLIQRKMTVIQITDDLFVNVNLRREVNTICE
KVEDAANGIIHMLIKGSESCNLFKSLKEMNYPLEFDLNGQSLDLSLGMNLTK
LIMDKINNEEDALILAGLKNLKMCLFHLHLDIGEMDYIVKSLISEPDLLEI
OIVSCCLANAVKILAIQNLHNLVRLSIDLSENYLEKNEALHELIDRMNVLEDTA
LMLPMGCVQVQSLSLKLEEVLPOLVAFKGNMLTDEIRIILGAFGKNLKNFQO
LNIAGNRVSDGMLAFMGVFEFLKQIVFDFSTKFLPLDPLVRLKLSKLTFLQE
ARLVGMQDDDDLSITGAFKLVTA"
BASE COUNT      436 a      248 c      327 g      384 t
ORIGIN

Query Match      26.7%: Score 836; DB 9; Length 1395;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2293 gtcgtctgactgacagctgtggtaacttgaagaaccttaagaagctcataatgataca 2352
|||||
DB 539 GTGGCTGACTGACAGCTGGGTAACCTTGAGAACCTTCAAAAGCTCAATATGATAACA 598

OY 2353 taaagatgaatgaagaagtgtctataaaactagctgaaggcctgaaaaacctgaagaaga 2412
|||||
DB 599 TAAAGATGATGAAGAAGATGCTATTAACCTGAGCTGAAGCGCTGAAAAACCTGAAGAAGA 658

OY 2413 tctgttattcatcttgaccacctgtctgacattgagaagggaattgattacatagta 2472
|||||
DB 659 TGTGTTATTATTATTTGACCCACTTGTCTGACATTGAGAGGGAATGATTCATATAGTCA 718

OY 2473 agctctgtcaagtgaaacctgtgacctgaagaattcaattagctcctgtctgt 2532
|||||
DB 719 ACTCTGCTGCAAGTGAACCCCTGACCTTGAGAAATTCATATAGCTCTGCTGCTGCTGT 778

Y 2533 ctgcaaatgcaatgaaaaatccctagctcagaatctcacaattgtgtcaaatgacattc 2592
|||||
DB 779 CTGCAAAATGCAATGAAAAATCTAGCTCAGCAATCTTCACAAATTTGGCAAACTGACGATTC 838

OY 2593 ttgatttaccagaataattcactggaagaagaatgagaatgaaactctctcagaactgacg 2652
|||||
DB 839 TTGATTTATCAGAAAATTATCTGGAATAAAGATGGAATTAAGCTCTTTCATGAACTGATCG 898

OY 2653 acaagatgaacgtgtctagaagacgtcaaccgcactgactgtcctcgaggctgtgaagtcg 2712
|||||
DB 899 ACAGATGAACTGCTAGAACAGCTCACCGCAGCTGATGCTGCCCTGCGGCTGTGACGCTGC 958

OY 2713 aaggaagcctggaagcctgttgaacaatttggaggaaggtcccaaacctcgtcaagcttg 2772
|||||
DB 959 AAGGACGCTGAGCAGCTGTTGAAACATTTGAGAGAGCTCCCAACACTGCAAGCTTG 1018

OY 2773 ggttcaaaaacggaagactcacagatacagaagattagaattttagttagattttttgga 2832
|||||
DB 1019 GGTGAAAAAACGAGACTCACAGATACAGATTAAGAAATTTTAGTGCAATTTTGGGA 1078

OY 2833 agaaccctctgaaaaaactccagagcttgaatttgcggaagaactcgtgtgaacagtgatg 2892
|||||
DB 1079 AGAACCTCTGAAAAACTCCAGAGTTGAATTTGGCGGGAATGCTGTGACAGAGTGAAG 1138

OY 2893 gatgctgtgctcgaatggtgtatttgaagaatccttaagaatagtglttttgaactta 2952
|||||
DB 1139 GATGCTGTGCTTCATGAGGTGTATTTGAGAAATCTTAAGCAATTAAGTGTTTTGAAGCTTGA 1198

OY 2953 gtactcaagaatcttactctgactcagcatagtcagaaaaaactagccaagtgatcatcca 3012

```

```

|||||
DB 1199 GTACTAAGAAATTTCTACCTGATCCAGACTTATGTCAGAAAACCTTACCAAGTCTATCA 1258

OY 3013 agttaaactttctgcaagaagcctgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3072
|||||
DB 1259 AGTTAACTTTTCTGCAAGAGAGCTAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1318

OY 3073 gttgtactcaggtgtctttaaactagtaactgtccttaataaagtgtactgaagc 3128
|||||
DB 1319 GTCTTATTTACAGGCTGCTTTTAAACTAGTACGCTTAAATTAAGTGTACTCGAAGC 1374

RESULT 9
AY027789 768 bp mRNA PRI 20-JUL-2001
LOCUS Homo sapiens CLAN1 (CLAN1) mRNA, complete cds.
DEFINITION AY027789
ACCESSION AY027789.1 GI:14324116
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE Clon, a novel human ced-4-like gene
JOURNAL Genomics. 75 (1-3), 77-83 (2001)
MEDLINE 21365712
PUBMED 11472070
REFERENCE 2 (bases 1 to 768)
AUTHORS Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source location/Qualifiers
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p22-p21"
/tissue_type="Lung"
1..768
/gene="CLAN1"
/gene="CLAN1"
/codon_start=1
/product="CLANB"
/protein_id="AAK14777.1"
/db_xref="GI:14324115"
/translation="MNFIKDNRSLIQRKMTVIQITDDLFVNVNLRREVNTICE
KVEDAANGIIHMLIKGSESCNLFKSLKEMNYPLEFDLNGQSLDLSLGMNLTK
LIMDKINNEEDALILAGLKNLKMCLFHLHLDIGEMDYIVKSLISEPDLLEI
OIVSCCLANAVKILAIQNLHNLVRLSIDLSENYLEKNEALHELIDRMNVLEDTA
LMLPMGCVQVQSLSLKLEEVLPOLVAFKGNMLTDEIRIILGAFGKNLKNFQO
LNIAGNRVSDGMLAFMGVFEFLKQIVFDFSTKFLPLDPLVRLKLSKLTFLQE
ARLVGMQDDDDLSITGAFKLVTA"
BASE COUNT 218 a 157 c 180 g 213 t
ORIGIN

Query Match 15.5%: Score 485; DB 9; Length 768;
Best Local Similarity 100.0%: Pred. No. 97e-251;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 gttggaagccttcatcagaacaatgaattcataaaggaacaatagccagcccttatt 71
|||||
DB 253 GTGGGAAGCTTTCATCCGAAACATGATTTCAATTAAGGACAAATGAGCCGACCTTATT 312

OY 72 caaagaatggaatgacgtgtataagaacaatacagaatgacctattgtatggaatgt 131
|||||
DB 313 CAAGAATGGAATGACGTGTATTAAGCAAAATCAAGATGACCTATTGTATGGAATGT 372

OY 132 ctgaatcgaaagaatgaacatcatctgctcgagaaggtgtagcagatgctgtctaga 191
|||||
DB 373 CTGAATCGGAAGAAGTAAACATCATTTTGCTGCGAAGAGGTGAGCAGAGATGCTCTAGA 432

```


QY 192 gggatcattacatgatttttgaagaggttcagagtccttgtaacccttcttcttaacc 251
 Db 433 gggatcattacatgatttttgaagaggttcagagtccttgtaacccttcttcttaacc 492
 QY 252 cttgaagagtggaactacatctcatttgaagacttgatgagacaaagcttcttctcag 311
 Db 493 cttgaagagtggaactacatctcatttgaagacttgatgagacaaagcttcttctcag 552
 QY 312 acatcagaagaagacttgagcagatttgctcagagatttgaagacttgatcaccatccca 371
 Db 553 acatcagaagaagacttgagcagatttgctcagagatttgaagacttgatcaccatccca 612
 QY 372 tcttctcgaacttatacccttgctggaagataatgacatttcttcttcttctgaagc 431
 Db 613 tcttctcgaacttatacccttgctggaagataatgacatttcttcttcttctgaagc 672
 QY 432 accctacagaacctgctcctgctggaagagaccacacacatcagcgcgtgagcagctg 491
 Db 673 accctacagaacctgctcctgctggaagagaccacacacatcagcgcgtgagcagctg 732
 QY 492 accct 496
 Db 733 accct 737

RESULT 10

LOCUS CNS01DS8/162692 bp DNA PRI 02-MAR-2000
 DEFINITION BAC sequence from the SP64 candidate region at 2p21-2p22 BAC 563N04
 of RPI-11 library from chromosome 2 of Homo sapiens (Human),
 complete sequence.

ACCESSION AL121658.2 GI:7159619
 VERSION AL121658
 KEYWORDS HTG: HTGS_DRAFT; SP64 genomic DNA interval.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 162692)
 Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Sanson, D.,
 Artiguenave, F., Davoine, C.S., Crisaud, C., Durr, A., Wincker, P.,
 Brothier, P., Catolico, L., Barbe, V., Burgeon, J.M.,
 Prud'Homme, J.F., Brice, A., Fontaine, B., Hellig, R. and
 Weissenbach, J.

TITLE Spastin, a novel AAA protein, is altered in the most frequent form
 of autosomal dominant spastic paraplegia
 JOURNAL Nat. Genet. (1999) in press
 REFERENCE 2 (bases 1 to 162692)
 Genoscope.

TITLE Direct Submission
 JOURNAL Submitted (29-FEB-2000) to the EMBL/GenBank/DBJ databases
 COMMENT On Mar 6, 2000 this sequence version replaced gi:6002391.
 FEATURES
 source Location/Qualifiers

1. 162692
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone_11b="RPI-11"
 /clone="563N04"
 BASE COUNT 44981 a 32017 c 32887 g 52792 t 15 others
 ORIGIN

Query Match 10.2% Score 318 DB 9: Length 162692:
 Best Local Similarity 100.0% Pred. No. 2.6e-160 Indels 0 Gaps 0
 Matches 318: Conservative 0 Mismatches 0

QY 2816 aggtgcatcttttgaagagaccccttgaaacatccacagcttgatggcgga 2875
 Db 152128 aggtgcatcttttgaagagaccccttgaaacatccacagcttgatggcgga 152159

QY 2876 tggctgagcagtgatgagtgctgctcctcatgggtgatattgagaatcttaaga 2935
 Db 152159 tggctgagcagtgatgagtgctgctcctcatgggtgatattgagaatcttaaga 2935

Db 152158 tggctgagcagtgatgagtgctgctcctcatgggtgatattgagaatcttaaga 152099
 QY 2936 aggtgcatcttttgaagagaccccttgaaacatccacagcttgatggcgga 2995
 Db 152098 aggtgcatcttttgaagagaccccttgaaacatccacagcttgatggcgga 152039
 QY 2996 tagcgaagtgtatcccaagtaaccttctcgaagaagcagccttgctgagcagc 3055
 Db 152038 tagcgaagtgtatcccaagtaaccttctcgaagaagcagccttgctgagcagc 151979
 QY 3056 tgatgatgatgatcagtgatgtattacaggttgctttaaactagtaactgataa 3115
 Db 151978 tgatgatgatgatcagtgatgtattacaggttgctttaaactagtaactgataa 151919
 QY 3116 ggtgactcgaagcagta 3133
 Db 151918 ggtgactcgaagcagta 151901

RESULT 11

LOCUS AC011232/185281 bp DNA HTG 10-MAR-2001
 DEFINITION Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
 7 unordered pieces.

ACCESSION AC011232.7 GI:13270720
 VERSION AC011232
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 185281)
 Waterston, R.H.
 The sequence of Homo sapiens clone
 unpublished
 JOURNAL 2 (bases 1 to 185281)
 Waterston, R.H.

REFERENCE Direct Submission
 JOURNAL Submitted (04-OCT-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Mar 10, 2001 this sequence version replaced gi:9799811.

COMMENT ----- Genome Center -----

Center: Washington University Genome Sequencing Center
 Center code: W0GSC
 Web site: http://genome.wustl.edu/gsc/index.shtml

Project Information
 Center project name: H.NH0078E13
 ----- Summary Statistics -----

Sequencing vector: M13, 57%
 Sequencing vector: plasmid, 40%
 Chemistry: Dye-terminator Big Dye, 50% of reads
 Assembly program: Phrap, version 0.990319
 Consensus quality: 182798 bases at least Q40
 Consensus quality: 183538 bases at least Q30
 Consensus quality: 184045 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 184681; sum-of-ctnigs
 Quality coverage: 7.07 in Q20 bases; sum-of-ctnigs
 Quality coverage: 7.33 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 7 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1985: contig of 1985 bp in length
 1986 2085: gap of unknown length

```

*      2086      8183: contig of 6098 bp in length
*      8184      8283: gap of unknown length
*      8284      22741: contig of 14458 bp in length
*      22742      22841: gap of unknown length
*      22842      41446: contig of 18605 bp in length
*      41447      41546: gap of unknown length
*      41547      69182: contig of 27636 bp in length
*      69183      69282: gap of unknown length
*      69283      112131: contig of 42849 bp in length
*      112132      112231: gap of unknown length
*      112232      185281: contig of 73050 bp in length.

```

FEATURES

```

Source
  1. .185281
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /clone="Rp11-78E13"
  1. .1985
    /note="assembly_name:Contig11"
  misc_feature
    /note="assembly_name:Contig12"
  misc_feature
    /note="assembly_name:Contig13"
  misc_feature
    /note="assembly_name:Contig14"
  misc_feature
    /note="assembly_name:Contig15"
  misc_feature
    /note="assembly_name:Contig16"
  misc_feature
    /note="assembly_name:Contig17"
BASE COUNT      52719 a 36957 c 37187 g 57816 t      602 others
ORIGIN

```

Query Match 10.2%: Score 318; DB 2; Length 185281;

Best Local Similarity 100.0%; Pred. No. 2.0e-160; Mismatches 0; Indels 0; Gaps 0;

Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      2816 aggtcatttttgaagaagaccctcgaagaactccagcagctgtaatttggcggaagaa 2875
      160183 AGGTCAATTTTGGAAAGAACCCCTGAAAACTTCCAGCGATTGCGGGGAAA 160124
      2876 tcgtgtgagcagtgatgagtgctgctcctcattggtgatttgaagaacttaagaact 2935
      160123 TCGTGTGAGCAGTGATGATGATGCTTGCCTCATGGGTGATTGAGAACTTAAGCAAT 160064
QY      2936 agtgttttgaacttaataaagaatttctaccctgataccagatcagtaagtaagaact 2995
      160063 AGTGTGTTTGAAGCTTAAGTAAGAAATTTCTACCTGATCCAGCATTAAGTCAAGAACT 160004
QY      2996 tagccaagtctatccaagtttaacttctcgaagaagcttagctgtgtgtgtgcaatt 3055
      160003 TAGCCAAGTCTTATCCAGATTACTTTCTGCAAGAACCTAGGCTTGTGGGTGCAATT 159944
QY      3056 tgatgatgatgatcagctgcttattacagtgctttaaactagtaactgcttaataaa 3115
      159943 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 159884
QY      3116 gtgtactcgaagccaagta 3133
      159883 GTGTACTCGAAGCCAGTA 159866

```

```

RESULT 12
LOCUS      AY027790      578 bp      mRNA
DEFINITION Homo sapiens CLAND (CLAN1) mRNA, complete cds.
ACCESSION      AY027790
VERSION      AY027790.1 GI:14324118
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE
AUTHORS      Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE      Clon, a novel human ced-4-like gene
JOURNAL      Genomics. 75 (1-3), 77-83 (2001)
MEDLINE      21365712
PUBMED      11472070
REFERENCE
AUTHORS      2 (bases 1 to 578)
TITLE      Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
JOURNAL      Direct Submission
SUBMITTED (21-FEB-2001) Program on Apoptosis and Cell Death
RESEARCH, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA

```

FEATURES

```

Source
  1. .578
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /map="2p22-p21"
    /tissue_type="Lung"
  1. .578
    /gene="CLAN1"
    /gene="CLAN1"
    /gene="CLAN1"
    /note="CARD protein"
    /codon_start=1
    /product="CLAND"
    /protein_id="AAK14779.1"
    /db_xref="GI:14324119"
    /translation="MNFIKDNRALLIGRMGTVIKQITDDLFWNVNIREVNITICE
      KVEDDARGIIMILKIGSESCNLELSLEKWNPLRQDINGOSILTA"
BASE COUNT      172 a 106 c 143 g 157 t
ORIGIN

```

Query Match 9.4%: Score 293; DB 9; Length 578;

Best Local Similarity 100.0%; Pred. No. 1e-146; Mismatches 0; Indels 0; Gaps 0;

Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      12 gttggaagctttccagaacaatgtaattcataaagacaatagcgagccctatt 71
      253 GGTGGAAAGCTTTCATCCAGAACAAATGAAATTCATTAAGAACCAATAGCCGACCTTAAT 312
QY      72 caaagaatgggaatgacgttataaagaacaatcagatgacctaattgtatggaatgt 131
      313 CAAGAATGGGAATGACGTATATAAGCAATCAACAGATGACCTATTGTATGGATGTT 372
QY      132 ctgaatcgagaagaataacatcattgtcgcgagaagtgagcagatgctgtaga 191
      373 CTGAATCGCGAAGAAGTAACATCATTTGCTGCGAGAAGTGAGAGAGATGCTGTAGA 432
QY      192 gggatcattcaatgattttgaaaaaggttcagagtcctgttaactcttcttaaac 251
      433 GGGATCATTCACATGATTTTGAAGAAAGGCTTCAGAGTCTCTTAACCTCTTTCAAAATCC 492
QY      252 cttaagagtggaactatcctctatttcaggaactggaatggaagaagcttt 304
      493 CTTAAGAGTGGAACTATCTCTATTTCAGGACTTGAAATGCAAAAGCTTTT 545

```

```

RESULT 13
LOCUS      G55568/c      553 bp      DNA
DEFINITION SHCC-100923 Human Homo sapiens SMS genomic, sequence tagged site.
ACCESSION      G55568
VERSION      G55568.1 GI:6120887
KEYWORDS      SMS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 553)

```



```
misc_feature /note="assembly_name:Contig9"
1181. 2476 /note="assembly_name:Contig12"
misc_feature /note="assembly_name:Contig13"
2577. 5035 /note="assembly_name:Contig14"
misc_feature 5136. 9606 /note="assembly_name:Contig15"
misc_feature 9707. 17078 /note="assembly_name:Contig16"
misc_feature 17179. 27158 /note="assembly_name:Contig17"
misc_feature 27259. 45137 /note="assembly_name:Contig18"
misc_feature 45238. 65522 /note="assembly_name:Contig19"
misc_feature 65623. 91498 /note="assembly_name:Contig20"
misc_feature 91599. 116835 /note="assembly_name:Contig21"
misc_feature 116936. 160583 /note="assembly_name:Contig22"
misc_feature 160583. 33703 g 47641 t 1047 others
BASE COUNT 44349 a 33843 c 33703 g 47641 t 1047 others
ORIGIN

Query Match 0.8% Score 25; DB 2; Length 160583;
Best Local Similarity 100.0%; Pred. NO. 0.15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ggtgggaagcttcacagacaa 36
Db 6875 GGTGGAGCTTTCATCCAGACAA 6851

RESULT 15
AC074195/c DNA HTG 20-AUG-2000
LOCUS Homo sapiens chromosome 11 clone RP11-750A9, WORKING DRAFT
DEFINITION AC074195 175152 bp
AC074195 175152 bp
AC074195.3 GI:9799883
AC074195.3 HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 175152)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 175152)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (16-JUL-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 13, 2000 this sequence version replaced gi:9743495.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0750A09
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
```

```
Assembly program: Phrap; version 0.990319
Consensus quality: 155181 bases at least Q40
Consensus quality: 161068 bases at least Q30
Consensus quality: 163835 bases at least Q20
Insert size: 174000; agarose-gel
Insert size: 172752; sum-of-contigs
Quality coverage: 4.16 in Q20 bases; agarose-gel
Quality coverage: 4.34 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1335: contig of 1335 bp in length
1336 1435: gap of unknown length
1436 3074: contig of 1639 bp in length
3075 3174: gap of unknown length
3175 5534: contig of 2360 bp in length
5535 5635: gap of unknown length
5636 8287: contig of 2653 bp in length
8288 8388: gap of unknown length
8389 10415: contig of 2028 bp in length
10416 10515: gap of unknown length
10516 12875: contig of 2360 bp in length
12876 12976: gap of unknown length
12977 16283: contig of 3308 bp in length
16284 16384: gap of unknown length
16385 19547: contig of 3163 bp in length
19548 24020: contig of 4374 bp in length
24021 24120: gap of unknown length
24121 28746: contig of 4626 bp in length
28747 28847: gap of unknown length
28848 33535: contig of 4689 bp in length
33536 33635: gap of unknown length
33636 38255: contig of 4620 bp in length
38256 38355: gap of unknown length
38356 43277: contig of 4922 bp in length
43278 43378: gap of unknown length
43379 47903: contig of 4526 bp in length
47904 52384: gap of unknown length
52385 52484: gap of unknown length
52485 58975: contig of 6491 bp in length
58976 59075: gap of unknown length
59076 67303: contig of 8228 bp in length
67304 67403: gap of unknown length
67404 74849: contig of 7446 bp in length
74850 74949: gap of unknown length
74950 83715: contig of 8666 bp in length
83716 83715: gap of unknown length
83716 94881: contig of 1116 bp in length
94882 94981: gap of unknown length
94982 104766: contig of 9785 bp in length
104767 104866: gap of unknown length
104867 118097: contig of 13231 bp in length
118098 118197: gap of unknown length
118198 132815: contig of 14618 bp in length
132816 132915: gap of unknown length
132916 148952: contig of 16037 bp in length
148953 149052: gap of unknown length
149053 175152: contig of 26100 bp in length.

FEATURES
source
1. 175152
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-750A9"
1. 1335
misc_feature
```

```

misc_feature      /note="assembly_name:Contig13"
                  1436..3074
                  /note="assembly_name:Contig16"
misc_feature      3175..5534
                  /note="assembly_name:Contig17"
misc_feature      5635..8287
                  /note="assembly_name:Contig18"
misc_feature      8388..10415
                  /note="assembly_name:Contig19"
misc_feature      10516..12875
                  /note="assembly_name:Contig20"
misc_feature      12976..16283
                  /note="assembly_name:Contig21"
misc_feature      16384..19546
                  /note="assembly_name:Contig22"
misc_feature      19647..24020
                  /note="assembly_name:Contig23"
misc_feature      24121..28746
                  /note="assembly_name:Contig24"
misc_feature      28847..33535
                  /note="assembly_name:Contig25"
misc_feature      33636..38255
                  /note="assembly_name:Contig26"
misc_feature      38356..43277
                  /note="assembly_name:Contig27"
misc_feature      43378..47903
                  /note="assembly_name:Contig28"
misc_feature      48004..52384
                  /note="assembly_name:Contig29"
misc_feature      52485..58975
                  /note="assembly_name:Contig30"
misc_feature      59076..67303
                  /note="assembly_name:Contig31"
misc_feature      67404..74849
                  /note="assembly_name:Contig32"
misc_feature      74950..83615
                  /note="assembly_name:Contig33"
misc_feature      83716..94881
                  /note="assembly_name:Contig34"
misc_feature      94982..104766
                  /note="assembly_name:Contig35"
misc_feature      104867..118097
                  /note="assembly_name:Contig36"
misc_feature      118198..132815
                  /note="assembly_name:Contig37"
misc_feature      132916..148952
                  /note="assembly_name:Contig38"
misc_feature      149053..175152
                  /note="assembly_name:Contig39"
BASE COUNT      42097 a 43906 c 45616 g 41104 t 2429 others
ORIGIN
Query Match      0.7%; Score 23; DB 2; Length 175152;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1922 agacacagtgatcacatg 1944
Db 80203 AGACACAGGTGATTCACATGG 80181

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
-----
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 183556)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 183556)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 15, 2000 this sequence version replaced gi:7684541.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0125F14
----- Summary Statistics -----
Sequencing vector: M13; 70x
Chemistry: Dye-primer ET; 70% of reads
Chemistry: Dye-terminator Big Dye; 30% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164779 bases at least Q40
Consensus quality: 170994 bases at least Q30
Consensus quality: 174798 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 180556; sum-of-ctrls
Quality coverage: 3.76 in Q20 bases; sum-of-ctrls
Quality coverage: 3.61 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1181: contig of 1181 bp in length
1182 1281: gap of unknown length
1282 2783: contig of 1502 bp in length
2784 2883: gap of unknown length
2883 4618: gap of 1735 bp in length
4619 4718: gap of unknown length
4719 6318: contig of 1600 bp in length
6319 6419: gap of unknown length
6419 8952: contig of 2534 bp in length
8953 9053: gap of unknown length
9053 11436: contig of 2384 bp in length
11437 11536: gap of unknown length
11537 14108: contig of 2572 bp in length
14109 14208: gap of unknown length
14209 16184: contig of 1976 bp in length
16185 16284: gap of unknown length
16285 20536: contig of 4252 bp in length
20537 20636: gap of unknown length
20637 25264: contig of 4628 bp in length
25265 25364: gap of unknown length
25365 29797: contig of 4433 bp in length
29798 29897: gap of unknown length
29898 33626: contig of 3729 bp in length
33627 33726: gap of unknown length
33727 37341: contig of 3615 bp in length
37342 37441: gap of unknown length
37442 41642: contig of 4201 bp in length
41643 41742: gap of unknown length
41743 47281: contig of 5549 bp in length
47292 47391: gap of unknown length
47392 50826: contig of 3435 bp in length
50827 50926: gap of unknown length

```

```

* 50927 58180: contig of 7254 bp in length
* 58181 58280: gap of unknown length
* 58281 63165: contig of 4885 bp in length
* 63166 63265: gap of unknown length
* 63266 68481: contig of 5216 bp in length
* 68482 73841: gap of unknown length
* 73842 73941: gap of unknown length
* 73942 79919: contig of 5978 bp in length
* 79920 80020: gap of unknown length
* 80020 86821: contig of 6702 bp in length
* 86822 92875: gap of unknown length
* 92876 92977: gap of unknown length
* 92977 98096: contig of 5120 bp in length
* 98097 98196: gap of unknown length
* 98197 106663: contig of 8467 bp in length
* 106664 106763: gap of unknown length
* 106764 117153: contig of 10390 bp in length
* 117154 117253: gap of unknown length
* 117254 128332: contig of 11079 bp in length
* 128333 128433: gap of unknown length
* 128433 139368: contig of 10936 bp in length
* 139369 139469: gap of unknown length
* 139469 154140: contig of 14672 bp in length
* 154141 154240: gap of unknown length
* 154241 167723: contig of 13483 bp in length
* 167724 167824: gap of unknown length
* 167824 183556: contig of 15733 bp in length.
Location/Qualifiers
1..183556
/organism="Homo sapiens"
/ds_xref="taxon:9606"
/chromosome="11"
/clone="RP11-125F14"
1..1181

```

```

misc_feature /note="assembly_name:Contig7"
misc_feature /note="assembly_name:Contig8"
misc_feature /note="assembly_name:Contig9"
misc_feature /note="assembly_name:Contig10"
misc_feature /note="assembly_name:Contig11"
misc_feature /note="assembly_name:Contig12"
misc_feature /note="assembly_name:Contig13"
misc_feature /note="assembly_name:Contig14"
misc_feature /note="assembly_name:Contig15"
misc_feature /note="assembly_name:Contig16"
misc_feature /note="assembly_name:Contig17"
misc_feature /note="assembly_name:Contig18"
misc_feature /note="assembly_name:Contig19"
misc_feature /note="assembly_name:Contig20"
misc_feature /note="assembly_name:Contig21"
misc_feature /note="assembly_name:Contig22"
misc_feature /note="assembly_name:Contig23"
misc_feature /note="assembly_name:Contig24"
misc_feature /note="assembly_name:Contig25"

```

```

misc_feature 68582..73841
/note="assembly_name:Contig26"
misc_feature 73942..79919
/note="assembly_name:Contig27"
misc_feature 80020..86721
/note="assembly_name:Contig28"
misc_feature 86822..92876
/note="assembly_name:Contig29"
misc_feature 92977..98096
/note="assembly_name:Contig30"
misc_feature 98197..106663
/note="assembly_name:Contig31"
misc_feature 106764..117153
/note="assembly_name:Contig32"
misc_feature 117254..128332
/note="assembly_name:Contig33
clone_end:SP6
vector_side:left"
128433..139368
/note="assembly_name:Contig34
clone_end:T7
vector_side:right"
139469..154140
/note="assembly_name:Contig35"
154241..167723
/note="assembly_name:Contig36"
167824..183556
/note="assembly_name:Contig37"
BASE COUNT 43865 a 46765 c 45344 g 44562 t 3020 others
ORIGIN

```

```

Query Match 0.7%; Score 23; DB 2; Length 183556;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1922 agacacagtgatccacatgg 1944
Db 103946 AGACACAGTGGATCCACATGG 103924
|||||

```

```

RESULT 17
AC090582 188459 bp DNA HTG 04-JUL-2001
LOCUS Homo sapiens chromosome 11 clone RP11-125F14 map 11, WORKING DRAFT
DEFINITION
SEQUENCE, 15 unordered pieces.
ACCESSION AC090582
VERSION AC090582.3 GI:14595831
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 188459)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-125F14
Unpublished
2 (bases 1 to 188459)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouhgatter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,D.S.,
Dodg,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galaan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Headford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Labocque,K., Lamazares,R., Landers,T.,
Lenczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Menes,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,

```


TITLE
JOURNAL
COMMENT

Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaddana, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Haas, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marcu, N., Mathews, C., McCarthy, M., McEwan, P., McKernan, K., McNeelers, R., Meldrum, J., Meneus, L., Mihova, T., Mienna, V., Murphy, T., Naylor, J., Nguyen, C., Noidu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Risse, C., Rogov, P., Roman, J., Rosett, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Triglilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 11, 2001 this sequence version replaced gi:14150930.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L11075

Center clone name: 750.H.9

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 199468 bases at least Q40
Consensus quality: 201173 bases at least Q30
Consensus quality: 202030 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 202987; sum-of-contigs
Quality coverage: 10.6 in Q20 bases; sum-of-contigs
Quality coverage: 10.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      450: contig of 450 bp in length
*      451 550: gap of 100 bp
*      551 3157: contig of 2607 bp in length
*      3158 3257: gap of 100 bp
*      3258 4660: contig of 1403 bp in length
*      4661 4760: gap of 100 bp
*      4761 6875: contig of 2115 bp in length
*      6876 6975: gap of 100 bp
*      6976 9750: contig of 2775 bp in length
*      9751 9850: gap of 100 bp
*      9851 12616: contig of 2766 bp in length
*      12617 12716: gap of 100 bp
*      12717 21088: contig of 8372 bp in length
*      21089 21188: gap of 100 bp
*      21189 27401: contig of 6213 bp in length
*      27402 27501: gap of 100 bp
*      27502 35375: contig of 7874 bp in length
*      35376 35475: gap of 100 bp
*      35476 77415: contig of 41940 bp in length
*      77416 77515: gap of 100 bp
*      77516 91045: contig of 13530 bp in length

```

```

*      91046 91145: gap of 100 bp
*      91146 113825: contig of 22680 bp in length
*      113826 113925: gap of 100 bp
*      113926 137980: contig of 24055 bp in length
*      137981 138080: gap of 100 bp
*      138081 158403: contig of 20323 bp in length
*      158404 158503: gap of 100 bp
*      158504 194539: contig of 36036 bp in length
*      194540 194639: gap of 100 bp
*      194640 204487: contig of 9848 bp in length.

```

FEATURES

source

```

1. 204487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-750H9"
/clone_lib="RC11 Human Male BAC"
1. 450
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
551. 3157
/note="assembly_fragment"
3258. 4660
/note="assembly_fragment"
4761. 6875
/note="assembly_fragment"
6976. 9750
/note="assembly_fragment"
9851. 12616
/note="assembly_fragment"
12717. 21088
/note="assembly_fragment"
21189. 27401
/note="assembly_fragment"
27502. 35375
/note="assembly_fragment"
35476. 77415
/note="assembly_fragment"
77516. 91045
/note="assembly_fragment"
91146. 113825
/note="assembly_fragment"
113926. 137980
/note="assembly_fragment"
138081. 158403
/note="assembly_fragment"
158504. 194539
/note="assembly_fragment"
194640. 204487
/note="assembly_fragment"
clone_end:r7
vector_side:right"

```

```

BASE COUNT      52691 a      48290 c      49218 g      52785 t      1503 others
ORIGIN

```

```

Query Match      0.7%: Score 23; DB 2; Length 204487;
Best Local Similarity 100.0%: Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1922 agacacagtggaatccacatgg 1944
|||||
DB 72357 AGACACAGTGTGAATCCACATGG 72379

```

```

RESULT 19
AP002957      87834 bp. DNA PRI 26-APR-2001
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:CTD-233717,
complete sequences.
ACCESSION AP002957

```


VERSION	AP002957.2	GI:13810522
KEYWORDS	HTG.	
SOURCE	Homo sapiens DNA, clone:CTD-233717.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (sites)	
TITLE	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.	
JOURNAL	Homo sapiens genomic DNA	
REFERENCE	Published Only in Database (2000) In press	
AUTHORS	2 (bases 1 to 87834)	
TITLE	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.	
JOURNAL	Direct Submission	
COMMENT	Submitted (28-NOV-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shuhiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail:hattori@gs.c.riken.go.jp, URL:http://hpg.gs.c.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170]	
FEATURES	On Apr 26, 2001 this sequence version replaced gi:11526584.	
SOURCE	Location/Qualifiers	
	1..87834	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="11"	
	/map="11q"	
	/clone="CTD-233717"	
BASE COUNT	23650 a 20616 c 20127 g 23441 t	
ORIGIN		
Query Match	0.7%; Score 22; DB 9; Length 87834;	
Best Local Similarity	100.0%; Pred. No. 6.3;	
Matches	22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1152 aaaacacacacacacataaag 1173	
Db	18515 AAAAACAAACACAAACATTAAG 18536	
RESULT 20		
AC020917		
LOCUS	AC020917 151088 bp DNA	PRI 08-NOV-2000
DEFINITION	Homo sapiens chromosome 19 clone CTD-2013N17, complete sequence.	
ACCESSION	AC020917	
VERSION	AC020917.4	GI:11120760
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 151088)	
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.	
JOURNAL	Direct Submission	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 151088)	
TITLE	DOE Joint Genome Institute.	
JOURNAL	Direct Submission	
COMMENT	Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
AUTHORS	3 (bases 1 to 151088)	
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.	
JOURNAL	Direct Submission	
COMMENT	Submitted (08-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
	On Nov 8, 2000 this sequence version replaced gi:7704958.	
	Draft Sequence Produced by DOE Joint Genome Institute	
	www.jgi.doe.gov	
	Finishing Completed at Stanford Human Genome Center	
	www.sngc.stanford.edu	
	Quality: Phrap Quality >40 99.8% of Sequence;	
	Estimated Total Number of Errors is 0.3.	

```

SRS Content:
SHGC-57513 G37313.
FEATURES
    source
        Location/Qualifiers
            1..151088
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="19"
                /clone="CID-2013N17"
BASE COUNT      40571 a 38233 c 38536 g 33748 t
ORIGIN
Query Match      0.7%: Score 22; DB 9; Length 151088;
Best Local Similarity 100.0%: Pident.No. 6.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1813 attactattgactcttga 1834
|||||
Db 117449 ATTACTATTGACTCTTTGA 117470

RESULT 21
AC007951
ID AC007951 standard: DNA; HTG: 151366 BP.
XX AC AC007951;
XX AC AC007951;
XX AC007951.2
DT 04-JUL-1999 (Rel. 60, Created)
DT 05-APR-2000 (Rel. 63, Last updated, Version 3)
XX Homo sapiens chromosome 9 clone RP11-99J1 map 9, WORKING DRAFT SEQUENCE, 9
DE unordered pieces.
XX
XX HTG: HTGS_DRAFT; HTGS_PHASE1.
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
XX [1]
RP 1-151366
RP Birren B., Linton L., Nusbaum C., Lander E.;
RT "Homo sapiens chromosome 9, clone RP11-99J1";
RL Unpublished.
XX
XX [2]
RP 1-151366
RP Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
RA Baker J., Baldwin J., Barua N., Beckert R., Benn J., Brown A., Castle A.,
RA Cerny J., Collangelo M., Collins S., Collamore A., Cooke P., DeArillano K.,
RA DeMayre E., Devon K., Dewar K., Donelan L., Doyle M., Ferreira P.,
RA Fitzhugh W., Forrest C., Funke R., Gage D., Galagan J., Gardyna S.,
RA Gilbert D., Grant G., Hagos B., Heatford A., Horton L., Howland J.C.,
RA Jones C., Kann L., Karatas A., Lehoczy J., Liu C., Locke K.,
RA MacDonald P., Margulis N., McMan P., McGuck A., McKernan K., McLaughlin J.,
RA Meadman J., Molla M., Morris W., Morrow J., Mychaleckyj J., Naylor J.,
RA Niloff M., O'Connor T., O'Donnell P., Pavlin B., Peterson K., Pollara V.,
RA Riley R., Roberts D., Roy A., Severy P., Stange-Thomann N., Stojanovic N.,
RA Stone C., Subramanian A., Testaye S., Torturella-Miller I., Vassiliev H.,
RA Vo A., Wagner A., Wheeler J., Wu X., Wyman D., Ye W.J., Zody M.;
Y
Submitted (02-JUL-1999) to the EMBL/GenBank/DBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
XX On Apr 1, 2000 this sequence version replaced g1.5332394.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC
CC Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research

```

CC	Center code: WIBR
CC	Web site: http://www.seq.wi.mit.edu
CC	Contact: sequence_submissions@genome.wi.mit.edu
CC	Project Information
CC	Center project name: L878
CC	Center clone name: 99_J_1
CC	----- Summary Statistics
CC	Sequencing vector: M13; M7815; 96% of reads
CC	Sequencing vector: Plasmid; n/a; 0.0% of reads
CC	3,80774032459426Chemistry: Dye-terminator-amershams; 96% of reads
CC	Chemistry: Dye-terminator Big Dye; 4% of reads
CC	Assembly program: Phrap; version 0.960731
CC	Consensus quality: 13664 bases at least Q40
CC	Consensus quality: 145968 bases at least Q30
CC	Consensus quality: 148942 bases at least Q20
CC	Insert size: 158000; agarose-gel
CC	Insert size: 150566; sum-of-contigs
CC	Quality coverage:
CC	* NOTE: This is a 'working draft' sequence. It currently
CC	* consists of 9 contigs. The true order of the pieces
CC	* is not known and their order in this sequence record is
CC	* arbitrary. Gaps between the contigs are represented as
CC	* runs of N, but the exact sizes of the gaps are unknown.
CC	* This record will be updated with the finished sequence
CC	* as soon as it is available and the accession number will
CC	* be preserved.
CC	* 1
CC	* 4870 4969: contig of 4869 bp in length
CC	* 4970 10428: contig of 5459 bp in length
CC	* 10429 10528: gap of 100 bp
CC	* 10529 18565: contig of 8037 bp in length
CC	* 18566 18665: gap of 100 bp
CC	* 18666 26259: contig of 7594 bp in length
CC	* 26260 26359: gap of 100 bp
CC	* 26360 41183: contig of 14824 bp in length
CC	* 41184 41283: gap of 100 bp
CC	* 41284 53957: contig of 12674 bp in length
CC	* 53958 54057: gap of 100 bp
CC	* 54058 81728: contig of 27671 bp in length
CC	* 81729 81828: gap of 100 bp
CC	* 81829 114786: contig of 32958 bp in length
CC	* 114787 114886: gap of 100 bp
CC	* 114887 151366: contig of 36480 bp in length.
XH	Key
FH	Location/Qualifiers
FT	source
FT	1. .151366
FT	/chromosome="g"
FT	/db_xref="taxon:9606"
FT	/organism="Homo sapiens"
FT	/map="g"
FT	/clone="RP11-99J1"
FT	/clone_lib="RPC1-11 Human Male BAC"
FT	1. .4869
FT	/note="assembly_fragment-clone_end:T7-vector_side:left"
FT	4970. 10428
FT	/note="assembly_fragment"
FT	10529. 18565
FT	/note="assembly_fragment"
FT	18666. 26259
FT	/note="assembly_fragment-clone_end:SP6-vector_side:right"
FT	26360. 41183
FT	/note="assembly_fragment"
FT	41284. 53957
FT	/note="assembly_fragment"
FT	54058. 81728
FT	/note="assembly_fragment"
FT	81829. 114786
FT	/note="assembly_fragment"
FT	114887. 151366
FT	/note="assembly_fragment"
XX	
Sequence	151366 BP; 46140 A; 28393 C; 28257 G; 47776 T; 800 other;

Query Match Similarity 0.7%; Score 22; DB 33; Length 151366;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1095 ttccactctcacacaacaa 1116
|||||
Db 132511 TTCCACTCTCACACAACAA 132532

RESULT 22
AC011980/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC011980 153733 bp DNA HTG 12-MAR-2000
Homo sapiens clone RP11-16H7, WORKING DRAFT SEQUENCE, 11 unordered
pieces.
AC011980
AC011980.3 GI:7230122
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 153733)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-16H7
Unpublished
2 (bases 1 to 153733)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domingo, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehocky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McMan, P., McGuck, A., McKernan, K., McLaughlin, J., Melidiri, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thoman, N., Stojanovic, N., Sudramanian, A., Talamas, J.,
Testaye, S., Tjirell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyma, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6453966.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L3543
Center clone name: 16.H.7

Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 117952 bases at least Q40
Consensus quality: 136478 bases at least Q30
Consensus quality: 146705 bases at least Q20
Insert size: 157000; agarose-1p
Insert size: 152733; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-1p
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 3262: contig of 3262 bp in length
* 1
* 3263 3362: gap of 100 bp
* 3363 9412: contig of 6050 bp in length
* 9413 9512: gap of 100 bp
* 9513 14144: contig of 4632 bp in length
* 14145 14244: gap of 100 bp
* 14245 24193: contig of 9949 bp in length
* 24194 24293: gap of 100 bp
* 24294 37092: contig of 12799 bp in length
* 37093 37192: gap of 100 bp
* 37193 49508: contig of 12316 bp in length
* 49509 49608: gap of 100 bp
* 49609 60861: contig of 11253 bp in length
* 60862 60961: gap of 100 bp
* 60962 76206: contig of 15245 bp in length
* 76207 76306: gap of 100 bp
* 76307 89151: contig of 12845 bp in length
* 89152 89251: gap of 100 bp
* 89252 107693: contig of 18442 bp in length
* 107694 107793: gap of 100 bp
* 107794 153733: contig of 45940 bp in length.
Location/Qualifiers
1.153733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC"
1.3262
/note="assembly-fragment"
misc_feature 3363..9412
/note="assembly-fragment"
misc_feature 9513..14144
/note="assembly-fragment"
misc_feature 14245..24193
/note="assembly-fragment"
misc_feature 24294..37092
/note="assembly-fragment"
misc_feature 37193..49508
/note="assembly-fragment"
misc_feature 49609..60861
/note="assembly-fragment"
misc_feature 49609..60861
/note="assembly-fragment"
clone_end:T7
vector_side:right"
/note="assembly-fragment"
60962..76206
/note="assembly-fragment"
76307..89151
/note="assembly-fragment"
clone_end:SP6
vector_side:right"
89252..107693
/note="assembly-fragment"
107794..153733
/note="assembly-fragment"

BASE COUNT 42218 a 34152 c 33610 g 42710 t 1043 others
ORIGIN

Query Match 0.7%: Score 22; DB 2; Length 153733;
Best Local Similarity 100.0%: Pred. No. 6.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1152 aaaaacacacacacataaag 1173
|||||
Db 120518 AAAACACACACACACATAAG 120497

RESULT 23
AL592213/c
LOCUS AL592213 155531 bp DNA HTG 04-AUG-2001

DEFINITION Homo sapiens chromosome 9 clone RP11-99J1, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL592213 AC007951
VERSION AL592213.6 GI:15131900
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 155531)
AUTHORS Hammond, S.
TITLE Direct Submissions
JOURNAL Submitted (03-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequest@sanger.ac.uk
On Aug 9, 2001 this sequence version replaced gi:15029558.
Draft Sequence produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: DA99J1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 30% of reads
Sequencing vector: plasmid; L08752; 69% of reads
Chemistry: Dye-terminator ET-amersham; 3% of reads
Chemistry: Dye-terminator Big Dye; 68% of reads
Dye-terminator Big Dye; 68% of reads
Chemistry: Dye-terminator-amersham; 28% of reads
Consensus quality: 155436 bases at least Q40
Consensus quality: 155453 bases at least Q30
Consensus quality: 155482 bases at least Q20
Insert size: 155531: sum-of-contigs
Insert size: 144681; 16.2% error; agarose-fp
Quality coverage: 15.49% in Q20 bases; sum-of-contigs Quality
coverage: 16.65% in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1.155531
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone_lib="RP11-99J1"
/clone_lib="RPCI-11.1"
1.155531
/note="assembly-fragment:03518
clone_end:SP6
vector_side:right"

BASE COUNT 50143 a 29768 c 28721 g 46899 t
ORIGIN

Query Match 0.7%: Score 22; DB 2; Length 155531;
Best Local Similarity 100.0%: Pred. No. 6.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1095 ttccactctcacacacacaa 1116
|||||
Db 38569 TTCCACCTCACACACACAA 38548

RESULT 24
AP002958/c
LOCUS AP002958 156230 bp DNA HTG 05-DEC-2000

Db 73708 AAAACAAACACACATTAAG 73687

|||||

RESULT 25
AC079194/C
LOCUS
DEFINITION
AC079194 159946 bp DNA HTG 22-NOV-2000
SEQUENCE, 37 unordered pieces.
AC079194.2 GI:11276162
HTG: HTGS_PHASE1, HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 159946)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaCocque,K.,
Lamazzari,R., Landers,T., Lechoczky,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mienga,Y.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnier,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (23-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2000 this sequence version replaced gi:9886030.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: I10601
Center clone name: 475.C.4

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 139893 bases at least Q40
Consensus quality: 150490 bases at least Q30
Consensus quality: 154263 bases at least Q20
Insert size: 17600; agarose-ftp
Insert size: 156346; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; agarose-ftp
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2597: contig of 2597 bp in length
2598 2697: gap of 100 bp
2698 3739: contig of 1042 bp in length
3740 3839: gap of 100 bp
3840 4797: contig of 958 bp in length
4798 4897: gap of 100 bp
4898 6197: contig of 1300 bp in length
6198 6297: gap of 100 bp
6298 8221: contig of 1924 bp in length
8222 8321: gap of 100 bp
8322 9558: contig of 1257 bp in length
9559 9658: gap of 100 bp
9659 11181: contig of 1523 bp in length
11182 11281: gap of 100 bp
11282 12622: contig of 1341 bp in length
12623 12722: gap of 100 bp
12723 14374: contig of 1652 bp in length
14375 14474: gap of 100 bp
14475 37771: contig of 23297 bp in length
37772 37871: gap of 100 bp
37872 39354: contig of 1483 bp in length
39355 39454: gap of 100 bp
39455 40854: contig of 1400 bp in length
40855 40954: gap of 100 bp
40955 42709: contig of 1755 bp in length
42710 42809: gap of 100 bp
42810 44167: contig of 1358 bp in length
44168 44267: gap of 100 bp
44268 46404: contig of 2137 bp in length
46405 46504: gap of 100 bp
46505 48310: contig of 1806 bp in length
48311 48410: gap of 100 bp
48411 51136: contig of 2726 bp in length
51137 51236: gap of 100 bp
51237 53831: contig of 2555 bp in length
53832 53931: gap of 100 bp
53932 56439: contig of 2508 bp in length
56440 56539: gap of 100 bp
56540 58998: contig of 2459 bp in length
58999 59098: gap of 100 bp
59099 63491: contig of 4393 bp in length
63492 63591: gap of 100 bp
63592 66743: contig of 3152 bp in length
66744 66843: gap of 100 bp
66844 69817: contig of 2974 bp in length
69818 69917: gap of 100 bp
69918 73103: contig of 3186 bp in length
73104 73203: gap of 100 bp
73204 77583: contig of 4380 bp in length
77584 77683: gap of 100 bp
77684 83310: contig of 5627 bp in length
83311 83410: gap of 100 bp
83411 87029: contig of 3619 bp in length
87030 87129: gap of 100 bp
87130 92085: contig of 4956 bp in length
92086 92185: gap of 100 bp
92186 97647: contig of 5462 bp in length
97648 97747: gap of 100 bp
97748 101467: contig of 3720 bp in length
101468 101567: gap of 100 bp
101568 106778: contig of 5211 bp in length
106779 106878: gap of 100 bp
106879 114040: contig of 7162 bp in length
114041 114140: gap of 100 bp
114141 123890: contig of 9750 bp in length
123891 123990: gap of 100 bp
123991 130670: contig of 6680 bp in length
130671 130770: gap of 100 bp
130771 139820: contig of 9050 bp in length
139821 139920: gap of 100 bp

```

*      139921 158959: contig of 19039 bp in length
*      158960 159059: gap of 100 bp
*      159060 159946: contig of 887 bp in length.
FEATURES
source
1. 159946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-475C4"
/clone_1lb="RPC1-11 Human Male BAC"
misc_feature
1. 2597
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
2698..3739
/note="assembly_fragment"
misc_feature
3840..4797
/note="assembly_fragment"
misc_feature
4898..6197
/note="assembly_fragment"
misc_feature
6298..8221
/note="assembly_fragment"
misc_feature
8322..9558
/note="assembly_fragment"
misc_feature
9659..11181
/note="assembly_fragment"
misc_feature
11282..12622
/note="assembly_fragment"
misc_feature
12723..14374
/note="assembly_fragment"
misc_feature
14475..37771
/note="assembly_fragment"
misc_feature
37872..39354
/note="assembly_fragment"
misc_feature
39455..40854
/note="assembly_fragment"
misc_feature
40955..42709
/note="assembly_fragment"
misc_feature
42810..44167
/note="assembly_fragment"
misc_feature
44268..46404
/note="assembly_fragment"
misc_feature
46505..48310
/note="assembly_fragment"
misc_feature
48411..51136
/note="assembly_fragment"
misc_feature
51237..53831
/note="assembly_fragment"
misc_feature
53932..56439
/note="assembly_fragment"
misc_feature
56540..58998
/note="assembly_fragment"
misc_feature
59099..63491
/note="assembly_fragment"
misc_feature
63592..66743
/note="assembly_fragment"
misc_feature
66844..69817
/note="assembly_fragment"
misc_feature
69918..73103
/note="assembly_fragment"
Query Match
Best Local Similarity 100.0%; Score 22; DB 2; Length 159946;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1152 aaaaacaacaactaag 1173
|||||
Db 150590 AAAACAACAACAATAAG 150569

```

```

RESULT 26
AL161913 186413 bp DNA HTG 11-JUL-2001
LOCUS

```

```

DEFINITION
Homo sapiens chromosome 9 clone RP11-64P11, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION
AL161913
VERSION
AL161913.10 GI:14329900
KEYWORDS
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 186413)
REFERENCE
Mashieghi-Mohammadi,M.
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14141268.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: B64P11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 7% of reads
Dye-terminator Big Dye; 92% of reads
Consensus quality: 185600 bases at least Q40
Consensus quality: 186182 bases at least Q30
Insert size: 186313; sum-of-contigs
Insert size: 140000; 37.9% error; agarose-fp
Quality coverage: 9.01x in Q20 bases; sum-of-contigs Quality
coverage: 12.46x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 168456: contig of 168456 bp in length
* 168457 168556: gap of 100 bp
* 168557 186413: contig of 17857 bp in length.
FEATURES
source
1. 186413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-64P11"
/clone_1lb="RPC1-11.1"
1. 168456
/note="assembly_fragment:00638"
misc_feature
168557..186413
vector_side:left"
clone_end:SP6
misc_feature
168557..186413
/note="assembly_fragment:00456"
BASE COUNT
60691 a 35149 c 33874 g 56599 t 100 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 22; DB 2; Length 186413;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1095 ttccacttcacacacaaca 1116
|||||
Db 174706 TTCACCTCTCACACACAACA 174685

```

```

RESULT 27
AL592438/c
LOCUS      AL592438      194533 bp      DNA
DEFINITION Homo sapiens chromosome 9 clone RP11-187G6, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION  AL592438
VERSION     AL592438.6
KEYWORDS    HTG: HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 194533)
AUTHORS     Wall,M.
TITLE       Direct Submission
JOURNAL     Submitted (01-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14787678.
COMMENT     ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba187G6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 9% of reads
Sequencing vector: plasmid; L08752; 90% of reads
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer-amersham; 9% of reads
Consensus quality: 194192 bases at least Q40
Consensus quality: 194424 bases at least Q30
Consensus quality: 194498 bases at least Q20
Insert size: 166314; sum-of-ctrligs
Insert size: 166314; agarose-fp
Quality coverage: 9.0ix in Q20 bases; sum-of-ctrligs Quality
coverage: 10.54x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..194533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-187G6"
/clone.lib="RPCT-11.1"
misc_feature
1..194533
/note="assembly-fragment:02981"
SE COUNT   61649 a 36707 c 36581 g 59596 t
IGIN
Query Match      0.7%; Score 22; DB 2; Length 194533;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1095 ttccactctcacacacaaacaa 1116
|||||
Db 194370 TTCACACTCTCACACACAAACA 194349
RESULT 28
BTCSK35
LOCUS      BTCSK35      7595 bp      DNA
DEFINITION Bovine gene for kappa-casein exons 3-5.
ACCESSION  X14908
VERSION     X14908.1
KEYWORDS    casein; kappa-casein; repetitive sequence; Alu-like repetitive

```

```

SOURCE      cow.
ORGANISM    Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 7595)
AUTHORS     Alexander,L.J., Stewart,A.F., Mackinlay,A.G., Kapelinskaya,T.V.,
Trach,T.M. and Gorodetsky,S.I.
TITLE       Isolation and characterization of the bovine kappa-casein gene
JOURNAL     Eur. J. Biochem. 178 (2), 395-401 (1988)
MEDLINE     89091174
REFERENCE   2 (bases 1 to 7595)
AUTHORS     Vaiman,D., Mercier,D., Moazami-Goudarzi,K., Eggen,A.,
Ciampolini,R., Lepingle,A., Velnala,R., Kuukinen,J., Varvio,S.L.,
Martin,P. et.al.
TITLE       A set of 99 cattle microsatellites: characterization, synteny
mapping, and polymorphism
JOURNAL     Mamm. Genome 5 (5), 288-297 (1994)
MEDLINE     94355772
REFERENCE   3 (bases 1 to 7595)
AUTHORS     Mackinlay,A.G.
TITLE       Direct Submission
JOURNAL     Submitted (16-MAR-1989) A.G. Mackinlay, The University of New South
Wales, P.O. Box 1 Kensington New South Wales Australia
location/Qualifiers
1..7595
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone.lib="EMBL3, EMBL4."
misc_feature
2886..2891
/label=sigstop
/usedin=X14907:kc_s1g
2886..2918
/label=ex3
/usedin=X14907:kc_cds
/usedin=X14906:kc_mrna
join(2892..2918,4930..5412)
/label=kc_mat
/product="kappa-casein"
4930..5412
/label=stop
/usedin=X14907:kc_cds
4930..5446
/label=ex4
/usedin=X14906:kc_mrna
5309
/note="c is t in kappa-casein B variant; changes acc (Thr)
to atc (Ile)"
5345
/note="a is c in kappa-casein B variant; changes gat (Asp)
to gct (Ala)"
5406
/note="a is g in kappa-casein B variant; loss of pslr
site"
5413
/note="a is t in kappa-casein B variant"
7296..7416
/label=ex5
/usedin=X14906:kc_mrna
7382
/note="t is c in kappa-casein B variant; loss of BglII
site, creates MspI site"
7416
7464
/note="t is c in variant clone"
7486
/note="t is c in variant clone"
7504
/note="t is c in variant clone"
7554
/note="g is a in variant clone"
7570..7571
variation

```

BASE COUNT 2646 a 1292 c 1182 g 2475 t
ORIGIN /note="c is inserted in variant clone"
Query Match 0.7%; Score 21; DB 4; Length 7595;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 415 ttcttaactgaaagcaccct 435
|||||
Db 6007 TTTTACTGAAAGCACCCT 6027
RESULT 29
LOCUS AC005199
DEFINITION Homo sapiens chromosome 16, cosmid clone 330D11 (LANL), complete
sequence.
ACCESSION AC005199
VERSION AC005199.1 GI:3273387
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 37906)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Mistra,M. and Deaven,L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 37906)
Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 37906)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Mistra,M. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source
1. 37906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="330D11"
/chromosome="16"
/map="16p13.3"
2548..2792
/rpt_family="MLT1"
4130..4301
/rpt_family="Alu"
5671..5997
/note="97% identity EST yb12f02.r1"
/db_xref="dbEST:T47043"
repeat_region
6373..6501
/rpt_family="MER25"
6572..6858
/rpt_family="Alu"
7074..7299
/rpt_family="L1"
7509..7828
/rpt_family="L1"
7898..10788
repeat_region
/rpt_family="L1"

repeat_region 11765..11854
/rpt_family="MER21"
repeat_region complement(12556..12682)
/rpt_family="MER2"
repeat_region complement(13150..13299)
/rpt_family="L1"
repeat_region 13754..14096
/rpt_family="Alu"
repeat_region 14119..14293
/rpt_family="MER5"
repeat_region complement(14450..14651)
/rpt_family="MER"
repeat_region 15660..15985
/rpt_family="Alu"
repeat_region complement(17523..17908)
/rpt_family="MSTa"
repeat_region 18375..18547
/rpt_family="MIR"
misc_feature complement(18614..18884)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region 19269..19921
/rpt_family="MER44C"
repeat_region 20666..27120
/rpt_family="L1"
repeat_region 22343..22577
/rpt_family="MER25"
repeat_region 30116..30259
/rpt_family="HSAT1"
repeat_region complement(30385..30580)
/rpt_family="HSAT1"
repeat_region complement(30678..30786)
/rpt_family="HSAT1"
repeat_region complement(31921..32340)
/rpt_family="MLT1"
repeat_region 33481..33615
/rpt_family="Alu"
repeat_region 33613..34213
/rpt_family="L1"
repeat_region 34664..34710
/rpt_family="MIR"
misc_feature complement(34811..35200)
/note="99% identity A0009624 BAC end sequence"
35732..36036
/rpt_family="Alu"
36016..36036
/note="(A)21"
repeat_region
/rpt_type=tandem
/rpt_unit=A
repeat_region complement(36219..36374)
/rpt_family="MLT1"
BASE COUNT 13076 a 7620 c 7679 g 9531 t
ORIGIN
Query Match 0.7%; Score 21; DB 9; Length 37906;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2012 caggactctgaggtcaccct 2032
|||||
Db 1512 CAGGACTCTGAGGTCCACT 1532
RESULT 30
LOCUS AC017948
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017948
VERSION AC017948.1 GI:6553242
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
TITLE
AUTHORS
JOURNAL

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 45496)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT

This sequence was identified as CDK:10212758 by the submitter. For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source
1..45496
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 13595 a 9140 c 9064 g 13697 t
ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 45496;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 29997 AAATCTGAAGCCATTAACA 29977
1666 aaattctgaagccataaca 1686
|||||

RESULT 31
LOCUS AC027055/c
DEFINITION Homo sapiens chromosome 3 clone RP11-334K8 map 3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC027055
VERSION AC027055.1 GI:7329416
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 76526)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-334K8
Unpublished
2 (bases 1 to 76526)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguski, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R., Melidim, J., Menus, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Norman, C.H., O'Connor, T., O'Donnell, P., Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thoman, N., Stojanovic, N., Subramanian, A., Talmas, J., Teste, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7376
Center clone name: 334_K_8

* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
804 903: gap of 100 bp
904 1690: contig of 787 bp in length
1691 1790: gap of 100 bp
1791 2559: contig of 769 bp in length
2560 2659: gap of 100 bp
2660 3445: contig of 786 bp in length
3446 3545: gap of 100 bp
3546 4348: contig of 803 bp in length
4349 4448: gap of 100 bp
4449 5243: contig of 795 bp in length
5244 5343: gap of 100 bp
5344 6119: contig of 776 bp in length
6120 6219: gap of 100 bp
6220 6998: contig of 779 bp in length
6999 7098: gap of 100 bp
7099 7884: contig of 786 bp in length
7885 7984: gap of 100 bp
7985 8766: contig of 782 bp in length
8767 8866: gap of 100 bp
8867 9669: contig of 803 bp in length
9670 9769: gap of 100 bp
9770 10559: contig of 790 bp in length
10560 10659: gap of 100 bp
10660 11449: contig of 790 bp in length
11450 11549: gap of 100 bp
11550 12330: contig of 781 bp in length
12331 12430: gap of 100 bp
12431 13222: contig of 792 bp in length
13223 13322: gap of 100 bp
13323 14098: contig of 776 bp in length
14099 14198: gap of 100 bp
14199 14992: contig of 794 bp in length
14993 15092: gap of 100 bp
15093 15881: contig of 789 bp in length
15882 15981: gap of 100 bp
15982 16771: contig of 790 bp in length
16772 16871: gap of 100 bp
16872 17655: contig of 784 bp in length
17656 17755: gap of 100 bp
17756 18549: contig of 794 bp in length
18550 18649: gap of 100 bp
18650 19440: contig of 791 bp in length
19441 19540: gap of 100 bp
19541 20335: contig of 795 bp in length
20336 20435: gap of 100 bp
20436 21236: contig of 801 bp in length
21237 21336: gap of 100 bp
21337 22140: contig of 804 bp in length
22141 22240: gap of 100 bp
22241 23041: contig of 801 bp in length

```

* 23042 23141: gap of 100 bp
* 23142 23942: contig of 801 bp in length
* 23943 24042: gap of 100 bp
* 24043 24040: contig of 798 bp in length
* 24841 24940: gap of 100 bp
* 24941 25729: contig of 789 bp in length
* 25730 25829: gap of 100 bp
* 25830 26603: contig of 774 bp in length
* 26604 26703: gap of 100 bp
* 26704 27480: contig of 777 bp in length
* 27481 27580: gap of 100 bp
* 27581 28381: contig of 801 bp in length
* 28382 28481: gap of 100 bp
* 28482 29279: contig of 798 bp in length
* 29280 29379: gap of 100 bp
* 29380 30167: contig of 788 bp in length
* 30168 30267: gap of 100 bp
* 30268 31056: contig of 789 bp in length
* 31057 31156: gap of 100 bp
* 31157 31955: contig of 799 bp in length
* 31956 32055: gap of 100 bp
* 32056 32844: contig of 789 bp in length
* 32845 32944: gap of 100 bp
* 32945 33731: contig of 787 bp in length
* 33732 33831: gap of 100 bp
* 33832 34582: contig of 751 bp in length
* 34583 34682: gap of 100 bp
* 34683 35480: contig of 798 bp in length
* 35481 35580: gap of 100 bp
* 35581 36381: contig of 801 bp in length
* 36382 36481: gap of 100 bp
* 36482 37276: contig of 795 bp in length
* 37277 37376: gap of 100 bp
* 37377 38182: contig of 806 bp in length
* 38183 38282: gap of 100 bp
* 38283 39072: contig of 790 bp in length
* 39073 39172: gap of 100 bp
* 39173 39983: contig of 811 bp in length
* 39984 40083: gap of 100 bp
* 40084 40873: contig of 780 bp in length
* 40874 40973: gap of 100 bp
* 40974 41759: contig of 786 bp in length
* 41760 41859: gap of 100 bp
* 41860 42652: contig of 793 bp in length
* 42653 42752: gap of 100 bp
* 42753 43560: contig of 808 bp in length
* 43561 43660: gap of 100 bp
* 43661 44448: contig of 788 bp in length
* 44449 44548: gap of 100 bp
* 44549 45335: contig of 787 bp in length
* 45336 45435: gap of 100 bp
* 45436 46220: contig of 785 bp in length
* 46221 46320: gap of 100 bp
* 46321 47108: contig of 788 bp in length
* 47109 47208: gap of 100 bp
* 47209 47983: contig of 775 bp in length
* 47984 48083: gap of 100 bp
* 48084 48863: contig of 780 bp in length
* 48864 48963: gap of 100 bp
* 48964 49755: contig of 792 bp in length
* 49756 49855: gap of 100 bp
* 49856 50659: contig of 804 bp in length
* 50660 50759: gap of 100 bp
* 50760 51542: contig of 783 bp in length
* 51543 51642: gap of 100 bp
* 51643 52427: contig of 785 bp in length
* 52428 52527: gap of 100 bp
* 52528 53328: contig of 801 bp in length
* 53329 53428: gap of 100 bp
* 53429 54211: contig of 783 bp in length
* 54212 54311: gap of 100 bp
* 54312 55088: contig of 777 bp in length
* 55089 55188: gap of 100 bp

```

```

Query Match      0.7%; Score 21; DB 2; Length 76526;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1146 atacagaaacaaacacaa 1166
Db 4183 ATACAGAAAACAAACACAAA 4163
|||||

```

```

RESULT 32
ATM4E13
LOCUS
DEFINITION
Arabidopsis thaliana DNA chromosome 4, BAC clone M4E13 (ESSAII
project).
ACCESSION
AL022023
VERSION
AL022023.1 GI:2924505
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 80346)
Bevan,M., Purnelle,B., Boutry,M., Goffeau,A., Hohelsel,J.,
Mewes,H.W., Mayer,K.F.X. and Schueller,C.
Unpublished
2 (bases 1 to 80346)
EU Arabidopsis sequencing/project.
Direct Submission
Submitted (04-APR-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk

```

```

FEATURES
source
1..80346
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(join(2..154,234..572))
/gene="M4E13.10"
2..572
/gene="M4E13.10"
complement(join(<2..154,234..572))
/gene="M4E13.10"

```

```

CDS
48964..49755: contig of 792 bp in length
49756..50659: contig of 804 bp in length
50660..50759: gap of 100 bp
50760..51542: contig of 783 bp in length
51543..51642: gap of 100 bp
51643..52427: contig of 785 bp in length
52428..52527: gap of 100 bp
52528..53328: contig of 801 bp in length
53329..53428: gap of 100 bp
53429..54211: contig of 783 bp in length
54212..54311: gap of 100 bp
54312..55088: contig of 777 bp in length
55089..55188: gap of 100 bp

```

exon

complement(2..154)

gene
/gene="M4E13.50"
/number=2
complement(join(20820..21014,21397..21499,21644..21723,
21807..21889,22003..22051,22140..22205,22752..22801,
22966..23090,23242..23354))
/gene="M4E13.60"
complement(join(20820..21014,21397..21499,21644..21723,
22966..23090,23242..23354))

CDS
Query Match 0.7%; Score 21; DB 8; Length 80346;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 caaatgctctctctctcgcg 646
|||||
db 75913 CAATTCGCTCTCTCTCGG 75933

RESULT 33

LOCUS ATT12J5 84499 bp DNA PLN 24-FEB-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T12J5 (ESSAII
project).
ACCESSION AL035522
VERSION AL035522.1 GI:4455339
KEYWORDS
SOURCE
ORGANISM
Chale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 84499)
Beyan,M., Murphy,G., Ridley,P., Hudson,S., Hobeisel,J., Mewes,H.W.,
Mayer,K.F.X. and Schueller,C.
Unpublished
2 (bases 1 to 84499)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (22-FEB-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Mike Beyan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.beyan@jbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosome 4 can be
viewed at: <http://webseq.mips.biochem.mpg.de/proj/thal/>.

FEATURES
source
1..84499
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
1..49380
/note="position 1-49380 overlaps to BAC clone M4E13, EMBL
acc:AL022023; for sequence analysis and annotation please
refer to this entry"

misc_feature
1..49380
/note="position 1-49380 overlaps to BAC clone M4E13, EMBL
acc:AL022023; for sequence analysis and annotation please
refer to this entry"

gene
48992..51335
/gene="T12J5.10"
complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/note="similarity to Human (H326) mRNA, Homo sapiens,
gb:U06631"
/codon_start=1
/product="putative protein"
/protein_id="CAB36721.1"
/db_xref="GI:4455340"
/translation="MSKARGRSGNGLHHVVDVEMREVGGISRNKRSFASSEN
LVRLIYKLEKKKGVNMYSPNAEDGLVLSGDDRRVYLMQDLQVKSFSISGHA
NNVQAKFMPSDRITVTCADGMFDLRTAPTELTFCRSVDPRRRMDATQLNATA

IDPNSNLFVAGMEVARYLDIRRFOEGINGFTRAADHFCPPHLIGNEDVGTGLA
FSEQSELVSNDEIYLFETPMGLGSPIDSSPIKSPVSKSESSSPKDEHSVS
LVYKGRKSEYKGVNFGPSEYVSGSDGRTITMRKGGELIRWEADRHVNCI
EPHPHIVPIASSGIESDIKVMTSKAFAPATIPENIELPSRPIPMISFLSRHVDDE
LFGNMDGIGIGNEBESIDESIDDAEDNDSDSYSSGVYLLDDNSDDDDSDDDVDMD
DSDDCFEIDNNNNNDGSGSETNVGSISSGSHDDDDDDDD"
complement(48992..49360)
/gene="T12J5.10"
/number=1
complement(49361..49761)
/number=1
complement(49762..50493)
/gene="T12J5.10"
/number=2
complement(50494..50814)
/number=2
complement(50815..51066)
/gene="T12J5.10"
/number=3
complement(51067..51206)
/number=3
complement(51207..51335)
/gene="T12J5.10"
/number=4
complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
complement(51806..52129)
/gene="T12J5.20"
/number=1
51806..53389
/gene="T12J5.20"
complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:U39301"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36722.1"
/db_xref="GI:4455341"
/translation="MEESKRNILDEEAKASLDIMRYVGFADIAAKCAIDKIPKAI
ENHSSOVLTELSIASVASPSHLRIIMRLVHOGLEKPTGDLATGYTNPLSR
RMKITKLHGKDLMAFADNLCIHSOLINAMACDARVYRPAACGGIFDVAIVYD
GGGGEITGILYKEPPIKGFNFDPHIEVAQVLDGVENEGMFDSPASDAVITK
WVLHDWGKDKCIKLIKCKEAVLPNIGVLIIECVIGKRTMTAEEDDKLEHVRQ
LDVMVMTSTGKERTLKMPFVLEAGFARYEVRDPOVSLIAYRS"
complement(52130..52610)
/number=1
complement(52611..52939)
/gene="T12J5.20"
/number=2
complement(52940..53064)
/number=2
complement(53065..53389)
/gene="T12J5.20"
/number=3
55847..57666
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/note="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:U39301
contains EST gb:N96576, W43582"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36723.1"
/db_xref="GI:4455342"
/translation="MSSDQLSKFLDRNKMEDNKKRVLDEEAKASLDIMKYVGFADIA
AAKCAIDKIPKAIENHSSOVLTELSIASVASPSHLRIIMRLVHOGLEKPTGDLATGYTNPLSR
DGLATGVNPMILSRIMTTRDGSLSAPFVETTPMLARLISVSVSSPVAGSIP
PPDAVAGKDVMSRQDNPI LSDMINEMACDARVYRPAACGHLGDFDGYTTVAVDG
GGTGTGMGLVKEPPIKGFNFDPHIEVAQVLDGVENEGMFDSPASDAVITK

VLHWDGDKCIKILKNCKEAVPPNIGKVLIVESVIGENKKTWIVDERDEKLEHVRML
DMVMAHTSTGTEKEMDFVLKEAGFARVEVDIDVOSLIIAYRS"
exon complement(55847. .56170)
/gene="t12j5.30"
/number=1
complement(56171. .56841)
/number=1
complement(56842. .57666)
/gene="t12j5.30"
/number=2
61697. .62188
/gene="t12j5.40"
61697. .62188
/gene="t12j5.40"
/number=1
61697. .62188
/gene="t12j5.40"
/note="similarity to predicted protein, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB36724.1"
/db_xref="GI:445343"
/translation="MTWPARSSPQNTKRPVYVHSPPTVDKISTGSGFSGSPL
NDQGVSNFQHSVAESSYSSGGLRNEYSVOVHDLDRTHDEDEYDMDGDEK
RRRIREFVSCLEFLVLAFTLECLILMGVSKSFATLTKVIDRDLRENFETISV
FKI"
gene complement(join(63375. .63860,63950. .64081,64151. .64551,
64736. .64869,64975. .65243))
/gene="t12j5.50"
complement(63375. .63860)
/gene="t12j5.50"
/number=1
63375. .65243
/gene="t12j5.50"
complement(join(63375. .63860,63950. .64081,64151. .64551,
64736. .64869,64975. .65243))
/gene="t12j5.50"
/note="strong similarity to lily mRNA, Lilium longiflorum,
gd:D21814
contains PTS HPR component phosphorylation sites
signatures, Pts_Hpr_Ser [SOEPMNLSRGVSLIG]"
/codon_start=1
/product="amino acid permease-like protein"
/protein_id="CAB36725.1"
/db_xref="GI:445344"
/translation="WSIALGNLFDESQSGSPLFMSPASTDPDPISGEKNGDGG
RIPVEMLPTESRGVNYTAFHLGSGIGVILPAPAFALGMVGTIILTVGFV
WLYTTLVLVQHEAVPGIIRISRYRLIASFGACTIIVIGKSIQOLIMSDDN
TAPLTSVOCELFVSCIAMISQPMNLSLFGVSLGAFGIAVCTIVITLIPVADSOR
TQVSYSATFMOKSFVIFNAIGLIALVYRGNLVEIOVLTNSOLKHDGTLPSDSK
NPSCTMRRAVMSIALVAICMPELFRVAYAYGCKIPATGSPYGNLYKLTIOESKR
NACFTHLFTFISCLSTYPTNLMPCADNEMVITTKRKPASTITVRMLVSLVCF
IAGVFPPLPYLAVLIGALALVLTFTYPCFMWISIKRPOKRSPLMFLNVLVGLASLS
VLLVASAMRLAQOKIHANFERP"
intron complement(63861. .63949)
/number=1
complement(63950. .64081)
/gene="t12j5.50"
/number=2
complement(64082. .64150)
/number=2
complement(64151. .64551)
/gene="t12j5.50"
/number=3
complement(64552. .64735)
/number=3
complement(64736. .64869)
/gene="t12j5.50"
/number=4
complement(64870. .64974)
/number=4
complement(64975. .65243)
exon

Query Match 0.7%: Score 21; DB 8; Length 84499;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 626 caaatcgtctctctcccg 646
|||||
Db 44958 CAAATTCGTCTCTCTCCG 44978
RESULT 34
ATF23E12
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
Unpublished
2 (bases 1 to 86710)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (01-APR-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
schneillemps.biochem.mpg.de,mayeremlps.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
FEATURES
source
gene
exon
CDS
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
173. .1890
/gene="F23E12.10"
173. .374
/gene="F23E12.10"
/number=1
join<173. .374,456. .919,1089. .1610,1678. .1890)
/gene="F23E12.10"
/codon_start=1
/product="hypothetical protein (fragment)"
/protein_id="CAA18726.1"
/db_xref="GI:3080407"
/db_xref="SPTREMBL:O65486"
/translation="ENAAVAGCFRAVESVLYMPSSSDSSTSLPNTTSWYLSAVTASV
VSEIGSGSEPAKVPDYPDERSPELPLVLLALGLVSLASRCTSSWTSADVSLNK
DAGIRPAVFPVGVGSLVGIILAVYEPVILWGNVDILKRPFPVKGSADLLDLVA
VKIATAMCRASGLVGYVAPSLFTGAGAMVYKFGIALAQNDNFNLSIEVASPO
AYGLGMATLAVGCOVPLTAVLLFELTODYRIYVPLIGAVGSMSTISGGSKQET
RETKETRRKKSQEAOSLTSSDESTNNLCVESSTCIDSLNSSEELPKSIFVSEA
MTRRAIVMSTSLBALTRMLTEKOSCALIVDPDITFGLITLSDIOFESKARKEGN
NRPKDIEVNDICSRSGKCKVPTVPMDDLAAOTINMKHLSHVAVVGSIIDAPRI
HPVGVLDRECIITLRR"
intron
exon
/gene="F23E12.10"
/number=1
456. .919
/gene="F23E12.10"
/number=2

```

intron      920..1088
            /gene="F23E12.10"
            /number=2
exon        1089..1610
            /gene="F23E12.10"
            /number=3
intron      1611..1677
            /gene="F23E12.10"
            /number=3
exon        1678..1890
            /gene="F23E12.10"
            /number=4
exon        3124..3241
            /gene="F23E12.20"
            /number=1
gene        3124..4428
            /gene="F23E12.20"
            /join(3124..3241,3323..3465,3595..3696,3945..4107,
            4211..4428)
            /note="Similarity to dihydroflavonol-4-reductase, Gerbera
            x sp., PIR2:S35189"
            /codon_start=1
            /product="putative protein"
            /protein_id="CA18727.1"
            /db_xref="GI:3080408"
            /db_xref="SPTREMBL:O65487"
            /translation="MDQAKGVCYTGASGLASMLVKRLLLEGVEVIGTVDPGNEK
            LAHLMLEGAKERLRLVKADLMERGSEPDNAIMGCGVHTASPEILRPATIGTLNVL
            SCRNPSLKRIVLTSSSTVWYALSKTLAEQAAKSESGCIDLTVTPSPFLVGPSP
            PDLGSTADVDGLKGETEKFQMGOMGVHIDVARHIVFEHEAAGKVICSSNV
            ISLEIVSFLSARVPSLPIPKRPDPLN"
            /number=1
intron      3242..3322
            /gene="F23E12.20"
            /number=1
exon        3323..3465
            /gene="F23E12.20"
            /number=2
intron      3466..3594
            /gene="F23E12.20"
            /number=2
exon        3595..3696
            /gene="F23E12.20"
            /number=3
intron      3697..3944
            /gene="F23E12.20"
            /number=3
exon        3945..4107
            /gene="F23E12.20"
            /number=4
intron      4108..4210
            /gene="F23E12.20"
            /number=4
exon        4211..4428
            /gene="F23E12.20"
            /number=5
gene        complement(join(4939..5013,5109..5195,5307..5342,
            5451..5651,5725..5811,6161..6163))
            /gene="F23E12.30"
            /number=1
gene        4939..6163
            /gene="F23E12.30"
            /complement(join(4939..5013,5109..5195,5307..5342,
            5451..5651,5725..5811,6161..6163))
            /gene="F23E12.30"
            /note="Contains Clathrin adaptor complexes small chain
            signature [VVKRYASLYF]
            contains EST gb:744219"
            /codon_start=1
            /product="clathrin assembly protein AP19 homolog"
            /protein_id="CA18728.1"
            /db_xref="GI:3080409"
            /db_xref="SPTREMBL:O23685"
            /translation="MIFVLLVSRKQGVRLTKWSPYAOKERKSVIRELSGVILNRP

```

```

exon        complement(4939..5013)
            /gene="F23E12.30"
            /number=1
intron      complement(5014..5108)
            /number=1
exon        complement(5109..5195)
            /gene="F23E12.30"
            /number=2
intron      complement(5196..5306)
            /number=2
exon        complement(5307..5342)
            /gene="F23E12.30"
            /number=3
intron      complement(5343..5450)
            /number=3
exon        complement(5451..5651)
            /gene="F23E12.30"
            /number=4
intron      complement(5652..5724)
            /number=4
exon        complement(5725..5811)
            /gene="F23E12.30"
            /number=5
intron      complement(5812..6160)
            /number=5
exon        complement(6161..6163)
            /gene="F23E12.30"
            /number=6
exon        complement(7352..7582)
            /gene="F23E12.40"
            /number=1
gene        7352..7582
            /gene="F23E12.40"
            /complement(7352..7582)
            /gene="F23E12.40"
            /complement(7352..7582)
            /gene="F23E12.40"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CA18729.1"
            /db_xref="GI:3080410"
            /db_xref="SPTREMBL:O65488"
            /translation="MPLVRKNGVMRIGSPMLHSENLVGDVEAPPGRDEIGEY
            GGETNVVSESGGEGGAQDAKTPPEGADRI"
            /number=1
gene        8233..9045
            /gene="F23E12.50"
            /complement(8233..9045)
            /gene="F23E12.50"
            /complement(8233..9045)
            /gene="F23E12.50"
            /number=1
CDS         complement(8233..9045)
            /gene="F23E12.50"
            /note="Similarity to DNA-binding protein PDI, Pisum
            sativum"
            /codon_start=1
            /product="putative protein"
            /protein_id="CA18730.1"
            /db_xref="GI:3080411"
            /db_xref="SPTREMBL:O65489"
            /translation="MELNRSEADAKAKAETTPTGATSSATASGSSGRPRPRPAGSK
            NKPKPTITRSDSPNVLRSHLEVTSQSDISEAVSTVATRRGCGCTIISGAVTNTV
            IROPAPAGGCVITLHGRFDLSTLTGTLALPPAPPGAGGLTVIAGGQGVGVGNVAG

```

```

Query Match      0.7%; Score 21; DB 8; Length 86710;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
07 626 caaatcgtctctctcccg 646
|||||

```

Db 11744 CAAATTCGCTCTCTCCTCCG 11764

```

RESULT 35
AC004907 130117 bp DNA PRI 21-AUG-1999
DEFINITION Homo sapiens clone DJ0853H20, complete sequence.
AC004907
AC004907.2 GI:5757545
HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 130117)
The sequence of Homo sapiens clone
Waterston,R.H.
REFERENCE
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 130117)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 130117)
Waterston,R.H.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 21, 1999 this sequence version replaced gi:3213090.
FEATURES
source 1..130117
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ0853H20"

BASE COUNT 43826 a 22237 c 22107 g 41947 t
ORIGIN

Query Match 0.7%; Score 21; DB 9; Length 130117;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2144 cctcagttgtctcagcac 2164
Db 28683 CCTCAGTTGTCTCCTCAGCAC 28703

RESULT 36
AC079351 157308 bp DNA HTG 03-FEB-2001
DEFINITION Homo sapiens chromosome 7 clone RP11-828B13, WORKING DRAFT
AC079351
AC079351.3 GI:12658193
HTG. HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 157308)
The sequence of Homo sapiens clone
Waterston,R.H.
REFERENCE
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 157308)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 3, 2001 this sequence version replaced gi:11415227.
COMMENT

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0828B13
----- Summary Statistics -----
Sequencing vector: MJ3; 93%
Sequencing vector: plasmid; 6%
Chemistry: Dye-Primer ET; 93% of reads
Chemistry: Dye-terminator Big Dye; 6% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151471 bases at least Q40
Consensus quality: 153455 bases at least Q30
Consensus quality: 154315 bases at least Q20
Insert size: 16200; agarose-fp
Insert size: 156208; sum-of-contigs
Quality coverage: 5.45 in Q20 bases; agarose-fp
Quality coverage: 5.28 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 6906: contig of 6906 bp in length
6907 7006: gap of unknown length
7007 16304: contig of 9298 bp in length
16305 16404: gap of unknown length
16405 27392: contig of 10988 bp in length
27393 27492: gap of unknown length
27493 38821: contig of 11359 bp in length
38822 38921: gap of unknown length
38922 53584: contig of 14662 bp in length
53584 53683: gap of unknown length
53683 71646: gap of 17963 bp in length
71646 71746: gap of unknown length
71746 96546: contig of 24800 bp in length
96547 96646: gap of unknown length
96647 119634: contig of 22988 bp in length
119635 119734: gap of unknown length
119735 146540: contig of 26886 bp in length
146541 146640: gap of unknown length
146641 150320: contig of 3680 bp in length
150321 150420: gap of unknown length
150421 153662: contig of 3242 bp in length
153663 153763: gap of unknown length
153763 157308: contig of 3546 bp in length.
FEATURES
source 1..157308
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-828B13"
1..6906
/note="assembly_name:Contig10"
7007..16304
/note="assembly_name:Contig11"
16405..27392
/note="assembly_name:Contig12"
27493..38821
/note="assembly_name:Contig13"
38922..53583
/note="assembly_name:Contig14
clone_end:T7
vector_side:left"
53684..71646
/note="assembly_name:Contig15"
71747..96546
/note="assembly_name:Contig16"

```

```

misc_feature      96647..119634
                  /note="assembly_name:Contig17"
misc_feature      119735..146540
                  /note="assembly_name:Contig18"
misc_feature      146641..150320
                  /note="assembly_name:Contig7"
misc_feature      150421..153662
                  /note="assembly_name:Contig8"
misc_feature      153763..157308
                  /note="assembly_name:Contig9
                  clone_end:SP6
                  vector_side:right"

BASE COUNT      52313 a 26687 c 27184 g 50018 t 1106 others
ORIGIN

Query Match      0.7% Score 21; DB 2; Length 157308;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2144 cctcagttgtgctcagcgc 2164
      |||||
Db 24594 CCTCAGTTGTCTCAGCAGC 24614

RESULT 37
AC024590/c      183839 bp DNA HTG 25-APR-2001
LOCUS           Homo sapiens chromosome 16 clone RP11-487C14, WORKING DRAFT
DEFINITION      AC024590
ACCESSION       AC024590
VERSION         AC024590.4 GI:13786378
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 183839)
AUTHORS         DOE Joint Genome Institute.
TITLE           Sequencing of Human Chromosome 16
JOURNAL         Unpublished
AUTHORS         2 (bases 1 to 183839)
REFERENCE       DOE Joint Genome Institute.
TITLE           Direct Submission
JOURNAL         Submitted (29-FEB-2000) Production Sequencing Facility, DOE Joint
AUTHORS         Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE           On Apr 25, 2001 this sequence version replaced gi:954669.
JOURNAL         -----Genome Center
AUTHORS         Center: Joint Genome Institute
TITLE           Center Code: JGI
JOURNAL         Web site: http://www.jgi.doe.gov
AUTHORS         -----
TITLE           Project Information
JOURNAL         Center Project Name: 593435
AUTHORS         Center clone name: RPCI-11_487C14
TITLE           -----
JOURNAL         Summary Statistics
AUTHORS         Consensus quality: 172462 bases at least Q40
JOURNAL         Consensus quality: 177746 bases at least Q30
AUTHORS         Consensus quality: 178816 bases at least Q20
JOURNAL         Estimated insert size: 163610; agarose-ftp estimation
AUTHORS         Estimated insert size: 182539; sum-of-contigs estimation
JOURNAL         Quality coverage: 14.49 in Q20 bases; agarose-ftp estimation
AUTHORS         Quality coverage: 12.99 in Q20 bases; sum-of-contigs estimation
JOURNAL         NOTE: This is a 'working draft' sequence. It currently
AUTHORS         consists of 14 contigs. The true order of the pieces
JOURNAL         is not known and their order in this sequence record is
AUTHORS         arbitrary. Gaps between the contigs are represented as
JOURNAL         * runs of N, but the exact sizes of the gaps are unknown.
JOURNAL         * This record will be updated with the finished sequence
JOURNAL         * as soon as it is available and the accession number will
JOURNAL         * be preserved.
      1
      1928: contig of 1928 bp in length

```

```

FEATURES
source
1..183839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-487C14"
/clone_lib="RPCI human BAC library 11"
BASE COUNT      50219 a 37425 c 39803 g 55084 t 1308 others
ORIGIN

Query Match      0.7% Score 21; DB 2; Length 183839;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 caggactctgaggtcacact 2032
      |||||
Db 42601 CAGGACTCTGAGGTCACACT 42581

RESULT 38
AC013553/c      184864 bp DNA HTG 13-DEC-2000
LOCUS           Homo sapiens chromosome 15 clone RP11-325L12 map 15, *** SEQUENCING
DEFINITION      AC013553
ACCESSION       AC013553
VERSION         AC013553.14 GI:11693406
KEYWORDS       HTG; HTGS_PHASE2.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 184864)
AUTHORS         Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE           Homo sapiens chromosome 15, clone RP11-325L12
JOURNAL         Unpublished
AUTHORS         2 (bases 1 to 184864)
JOURNAL         Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
AUTHORS         Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
AUTHORS         Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
AUTHORS         Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
AUTHORS         Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
AUTHORS         Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
AUTHORS         Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
AUTHORS         Lebecky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
AUTHORS         McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
AUTHORS         Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

```


Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessier,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 13, 2000 this sequence version replaced gi:9280765.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L2203
Center clone name: 325_L_12

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 184864: contig of 184864 bp in length.
Location/Qualifiers
1. 184864

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-325L12"
/clone_lib="RPC1-11 Human Male BAC"

BASE COUNT 51274 a 42968 c 41143 g 49479 t
ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 184864;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 294 caaagcttttcacagaca 314
|||||
DB 46943 CAAAGCTTTTCATCAGACA 46923

RESULT 39

AC080090 194874 bp DNA HTG 26-NOV-2000
Homo sapiens chromosome 7 clone RP11-79616, WORKING DRAFT SEQUENCE,
29 unordered pieces.

AC080090
AC080090.3 GI:11345029
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
SOURCE
ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 194874)

Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 194874)

Waterston,R.H.
Direct Submission
Submitted (23-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

On Nov 26, 2000 this sequence version replaced gi:11276326.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0796106
----- Summary Statistics -----

Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184228 bases at least Q40
Consensus quality: 186326 bases at least Q30
Insert size: 202000; agarose-fp
Insert size: 191473; sum-of-contigs
Quality coverage: 3.95 in Q20 bases; agarose-fp
Quality coverage: 4.25 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 601: contig of 601 bp in length
1 701: gap of unknown length
* 602 701: contig of 1832 bp in length
* 2533 2633: gap of unknown length
* 2534 2634: gap of unknown length
* 2634 3924: contig of 1291 bp in length
* 3925 4024: gap of unknown length
* 4025 5276: contig of 1252 bp in length
* 5277 5376: gap of unknown length
* 5377 6812: contig of 1436 bp in length
* 6813 6912: gap of unknown length
* 6913 9058: contig of 2146 bp in length
* 9059 9158: gap of unknown length
* 9159 11288: contig of 2130 bp in length
* 11289 11388: gap of unknown length
* 11389 13955: contig of 2567 bp in length
* 13956 14055: gap of unknown length
* 14056 16269: contig of 2214 bp in length
* 16270 16370: gap of unknown length
* 16371 19135: contig of 2765 bp in length
* 19136 19234: gap of unknown length
* 19235 22939: contig of 3705 bp in length
* 22940 23039: gap of unknown length
* 23040 26095: contig of 3055 bp in length
* 26096 26195: gap of unknown length
* 26196 29786: contig of 3591 bp in length
* 29787 29886: gap of unknown length
* 29887 32815: contig of 2928 bp in length
* 32816 32915: gap of unknown length
* 32916 37640: contig of 4723 bp in length
* 37641 41001: gap of unknown length
* 41002 41101: contig of 3261 bp in length
* 41102 46784: gap of unknown length
* 46785 46884: contig of 5684 bp in length
* 46885 54478: gap of unknown length
* 54479 54578: gap of unknown length
* 54579 61717: contig of 7138 bp in length
* 61718 61817: gap of unknown length
* 61818 68787: contig of 6970 bp in length
* 68788 68887: gap of unknown length
* 68888 76504: contig of 7618 bp in length
* 76505 86157: gap of unknown length
* 86158 86157: contig of 9553 bp in length

* 86158 86257: gap of unknown length
* 86258 95154: contig of 8897 bp in length
* 95155 95254: gap of unknown length
* 95255 105637: contig of 10383 bp in length
* 105638 105737: gap of unknown length
* 105738 119880: contig of 14143 bp in length
* 119881 119980: gap of unknown length
* 119981 131202: contig of 11222 bp in length
* 131203 131302: gap of unknown length
* 131303 144326: contig of 13024 bp in length
* 144327 144426: gap of unknown length
* 144427 164332: contig of 19906 bp in length
* 164333 164433: gap of unknown length
* 164433 194874: contig of 30442 bp in length.
* 194874
Location/Qualifiers
1. 194874
/organism="Homo Sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-79616"
1. .601
misc_feature /note="assembly_name:Contig10"
702. 2533
/note="assembly_name:Contig19"
2634. .3924
misc_feature /note="assembly_name:Contig11"
4025. .5276
/note="assembly_name:Contig14"
5377. .6812
misc_feature /note="assembly_name:Contig16"
6913. .9058
/note="assembly_name:Contig17"
9159. .11288
misc_feature /note="assembly_name:Contig18"
11389. .13955
/note="assembly_name:Contig20"
14056. .16269
misc_feature /note="assembly_name:Contig21"
16370. .19134
/note="assembly_name:Contig22"
19235. .22939
misc_feature /note="assembly_name:Contig23"
23040. .26095
/note="assembly_name:Contig24"
26196. .29786
misc_feature /note="assembly_name:Contig25"
29887. .32814
/note="assembly_name:Contig26"
32915. .37639
misc_feature /note="assembly_name:Contig27"
37740. .41000
/note="assembly_name:Contig28"
41101. .46784
misc_feature /note="assembly_name:Contig29"
46885. .54478
/note="assembly_name:Contig30"
54579. .61716
misc_feature /note="assembly_name:Contig31"
61817. .68786
/note="assembly_name:Contig32"
68887. .76504
misc_feature /note="assembly_name:Contig33"
76605. .86157
/note="assembly_name:Contig34"
86258. .95154
misc_feature /note="assembly_name:Contig35"
95255. .105637
/note="assembly_name:Contig36"
105738. .119880
misc_feature /note="assembly_name:Contig37"
119981. .131202
/note="assembly_name:Contig38"
131303. .144326
misc_feature

misc_feature /note="assembly_name:Contig39"
14427. .164332
/note="assembly_name:Contig40"
misc_feature 164433. .194874
/note="assembly_name:Contig41"
BASE COUNT 63657 a 32523 c 33628 g 62260 t 2806 others
ORIGIN
Query Match 0.7%; Score 21; DB 2; Length 194874;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2144 cctcattgtgcctcagcac 2164
|||||
Db 52703 CCTCAGTTGGTCCTCAGCAC 52723
|||||
RESULT 40
ATCHRIV82
LOCUS DEFINITION
ACCESSION ALL61586
VERSION
KEYWORDS
SOURCE ORGANISM
Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 29459 to 132608; 129923 to 195165)
Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
2 (bases 98968 to 179313)
Purnelle,B., Boutry,M., Goffeau,A., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
3 (bases 1 to 44270)
Terryn,N., Ardijns,W., Buysseart,C., Dasseville,R., de Clerck,R.,
De Keyser,A., Neyt,P., Rouze,P., Van Den Daele,H., Villarroel,R.,
Glehen,J., Van Montagu,M., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
Unpublished
4 (bases 1 to 195165)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@jbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV81 at the 5' end and an
overlap with ATCHRIV83 at the 3' end.
Location/Qualifiers
1. 195165
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
8478. .8759
exon /gene="AT4g34680"
/number=1
8478. .9388
/gene="AT4g34680"
join(8478. .8759, 8861. .9388)
/gene="AT4g34680"
/note="contains EST gb:AI994545.1, R30028"
/codon_start=1
/product="GATA transcription factor 3"

/protein_id="CAB80185.1"
 /db_xref="GI:7270419"
 /translation="MEIATEARLAKSLRGESTISLKHQIVASEDLISRTSSLPEDFS
 VECFLDFSGQEEEEVSVSSSOEEOEDHCVSOPCFIDLPSPDDVELE
 WYSRVDDCSSPEVSLTLTQHKTPSPSRILEVPRTRSRNSLTGSRWPLVSTNHQ
 HAATQDLRRKKQETLVFQRCSHCGTNTPDWRTGVPYGPKTLCNACGVRFFSGRLCP
 EYRPADSPETSNEIHSNLHRKVLRLRKSKEIGEETGEASTKSDPVKFGSKW"
 8760. .8860
 /gene="AT4g34680"
 /number=1
 8861. .9388
 /gene="AT4g34680"
 /number=2
 10241. .10875
 /gene="AT4g34690"
 complement(10241. .10855)
 /gene="AT4g34690"
 /number=1
 complement(join(10241. .10855,10867. .10875))
 /gene="AT4g34690"
 complement(join(10241. .10855,10867. .10875))
 /gene="AT4g34690"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB80186.1"
 /db_xref="GI:7270420"
 /translation="MDLSDDGSGEKKKSDSGDKKSDGSSNNDDDDYTPNDPOS
 TWFKYFLATRLHYGVKRCSESVKRWKFLKTNISQAKLQLLAKAKKEHOTF
 VEYTLGKASRDSQTYSTFLTWCFGSJLTCSFVWGLMNTSKTRSLADMLEINKK
 SVYIHEFFKEMSAKNALYMRIDETITKAKONQEDLOKIVALLVSKK"
 complement(10856. .10866)
 /number=1
 complement(10867. .10875)
 /gene="AT4g34690"
 /number=2
 11652. .13140
 /gene="AT4g34700"
 join(11652. .11783,12719. .12800,12946. .13001,13099. .13140)
 /gene="AT4g34700"
 /note="contains EST gb:T04342, T13747, AA040992,
 A1992559.1, Z33673, H36271, T43433, T43941"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80187.1"
 /db_xref="GI:7270421"
 /translation="MSGVSTAYFARRAQAQKRVVILYRRAKDTLNNAAVHHRIYRD
 EDVDRIDKLIAHGEAEYKMKRHPDPIVPMAGSGKFCENPTPPAGIETVNYGLEDN
 P"
 11652. .11783
 /gene="AT4g34700"
 /number=1
 11784. .12718
 /gene="AT4g34700"
 /number=1
 12719. .12800
 /gene="AT4g34700"
 /number=2
 12801. .12945
 /gene="AT4g34700"
 /number=2
 12946. .13001
 /gene="AT4g34700"
 /number=3
 13002. .13098
 /gene="AT4g34700"
 /number=3
 13099. .13140
 /gene="AT4g34700"
 /number=4
 complement(15093. .17228)
 /gene="AT4g34710"
 complement(15093. .17228)
 /gene="AT4g34710"
 /translation="MAPVNLRTNLSVRRTRVROMANPRRVKVAQKIMRELSMDLITDT

/note="Contains Orn/DAP/Arg decarboxylases family 2
 signatures AA144-162:Orn/DAP/Arg decarboxylases family 2
 signatures AA325-338
 contains EST gb:W43783, Z33699, Z33688, T46784,
 A1998693.1, AA404845"
 /codon_start=1
 /product="arginine decarboxylase SPE2"
 /protein_id="CAB80188.1"
 /db_xref="GI:7270422"
 /translation="MAPALACVDTSTVPVPAFAFSDRAGVDFITPASSFTSANVYVDRMS
 SLSSSLYRIDGWGAPYFIANSNGNISVPRHGETLPDQDIDLKTVKRYTPKSSGGL
 GLQPLIVAFEPVLNRLCLOASDAPDYKSGYSDYVPCNODRFEVEDIVK
 FGSSEFPGLEASKEPILAMSCLCGSPDALVNGFXDAEYISLALGKRLANTV
 IYLEOEELDIVETLSOKMNVPRVIGLRKRLKTKSGHSGSGRGRGLTTOIVR
 VYRKLRQSGMDLCLDLHFIHSQITPSTSLSDGYAEAAQITCELYRGAHKVYDIDG
 GGLGIDYDGSKGSSEDSLVAVSLSEYAEVAVSVVCDRSSVKHPVTCSEGRAIVS
 HHSVLIFEAVSADKPMVQATPGDIOFLLEGEEARANEEDLYAAVMRMDHSCILY
 DOLKORCVGEKEGVLSIEOLASVDGLSEYKAIGASPPVPTVNIINLSVPTSIDLW
 GIDQLFPIYPIHKLDORPGARCIISDLTCDSDGKINKFTGSGSSLPLEHLDNGSGR
 YELGMEIGAYEALGVYHNLFGSPSVYRVSQDGPSPHSAVTRAVPGGSSADVLRAMO
 HPEPLMFOTIKRAEEMHTKGSSEGENEEDEDFNNVAASLDPSRNMPLYLATEQ
 ASPNSLSAISNLGFTYCDQEDVYDTISA"
 complement(15093. .17228)
 /gene="AT4g34710"
 /number=1
 15093. .17228
 /gene="AT4g34710"
 complement(join(23001. .23133,23231. .23516,23868. .23943))
 /gene="AT4g34720"
 complement(23001. .23133)
 /gene="AT4g34720"
 /number=1
 23001. .23943
 /gene="AT4g34720"
 complement(join(23001. .23133,23231. .23516,23868. .23943))
 /gene="AT4g34720"
 /note="Contains Prokaryotic membrane lipoprotein lipid
 attachment site AA16-26:Prokaryotic membrane lipoprotein
 lipid attachment site AA91-101
 contains EST gb:T41774, N96995, N38619, Z26112, N38618,
 A1998496.1, A1996086.1, R90087, AA12128, H76763"
 /codon_start=1
 /product="vacuolar H⁺-transporting ATPase 16k chain"
 /protein_id="CAB80189.1"
 /db_xref="GI:7270423"
 /translation="MSTRSGDETFAPFPGFLGAAALVYSCGAATGTAKSGVAGSAMG
 VWRPELVKMSIVPVMAVGLIGLILVISTGINPAKSYLFDGVAHLSSGLACG
 LAGLSGMAIGIVGDAGVYRANAOQPKLEFGMLILIFAEALALYGLIVGILSSRAGQ
 SRAE"
 complement(23134. .23230)
 /number=1
 complement(23231. .23516)
 /gene="AT4g34720"
 /number=2
 complement(23517. .23867)
 /number=2
 complement(23868. .23943)
 /gene="AT4g34720"
 /number=3
 24797. .25947
 /gene="AT4g34730"
 complement(join(24797. .24955,25226. .25399,25678. .25947))
 /gene="AT4g34730"
 complement(join(24797. .24955,25226. .25399,25678. .25947))
 /gene="AT4g34730"
 /note="similarity to similarity predicted protein,
 Synechocystis sp., PIR2:576285
 contains EST gb:F19817"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80190.1"
 /db_xref="GI:7270424"
 /translation="MAPVNLRTNLSVRRTRVROMANPRRVKVAQKIMRELSMDLITDT

VL0HAYLPEALCADRYLSTTTSDVEVSNDLQWRRPNKKLCLNKKVYVSFGDD

Query Match

Best Local Similarity 0.7%; Score 21; DB 8; Length 195165;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 636 caattgctctctctccgcg 646

Db 174880 CAAATCTCTCTCTCCG 174900

RESULT 41

AC025566/c

DEFINITION

Homo sapiens chromosome 3 clone RP11-50102, WORKING DRAFT SEQUENCE,

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

AC025566

(see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 40 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

17628: contig of 17628 bp in length
17728: gap of unknown length
17729: contig of 12923 bp in length
30651: gap of unknown length
30751: gap of unknown length
30752: contig of 9059 bp in length
39811: gap of unknown length
39910: contig of 7994 bp in length
47904: gap of unknown length
47905: contig of 8306 bp in length
48005: gap of unknown length
56311: gap of unknown length
56411: contig of 7158 bp in length
63668: gap of unknown length
63669: contig of 6656 bp in length
70324: gap of unknown length
70325: gap of unknown length
70424: gap of unknown length
77997: contig of 7573 bp in length
78097: gap of unknown length
85133: contig of 7036 bp in length
85134: gap of unknown length
85234: contig of 5950 bp in length
91183: gap of unknown length
91283: contig of 4384 bp in length
95667: gap of unknown length
95668: contig of 5969 bp in length
101736: gap of unknown length
101836: gap of unknown length
101837: contig of 6802 bp in length
108639: gap of unknown length
108738: gap of unknown length
114324: contig of 5586 bp in length
114424: gap of unknown length
118727: contig of 4303 bp in length
118728: gap of unknown length
118827: contig of 3942 bp in length
122769: gap of unknown length
122869: contig of 6463 bp in length
129332: gap of unknown length
129333: contig of 4734 bp in length
129433: gap of unknown length
134167: contig of 5058 bp in length
134267: gap of unknown length
139424: contig of 4865 bp in length
139425: gap of unknown length
144289: contig of 4773 bp in length
144389: gap of unknown length
149162: contig of 3669 bp in length
149163: gap of unknown length
152931: contig of 4633 bp in length
152932: gap of unknown length
157664: contig of 5662 bp in length
157665: gap of unknown length
163426: contig of 3277 bp in length
163527: gap of unknown length
166803: contig of 2275 bp in length
166904: gap of unknown length
169178: contig of 3370 bp in length
169279: gap of unknown length
172648: contig of 2431 bp in length
172749: gap of unknown length
175179: contig of 2093 bp in length
175279: gap of unknown length
177372: contig of 1864 bp in length
177373: gap of unknown length
179336: contig of 1092 bp in length
179437: gap of unknown length
180528: contig of 1601 bp in length
180529: gap of unknown length
182228: contig of 1601 bp in length
182329: gap of unknown length

* NOTE: Estimated insert size may differ from sequence length

* 182330 184954: contig of 2625 bp in length
* 184955 185054: gap of unknown length
* 185055 186495: contig of 1441 bp in length
* 186496 186595: gap of unknown length
* 186596 188317: contig of 1722 bp in length
* 188318 188417: gap of unknown length
* 188418 189682: contig of 1265 bp in length
* 189683 189782: gap of unknown length
* 189783 191715: contig of 1933 bp in length
* 191716 191815: gap of unknown length
* 191816 192922: contig of 1107 bp in length
* 192923 193022: gap of unknown length
* 193023 194102: contig of 1080 bp in length
* 194103 194202: gap of unknown length
* 194203 195217: contig of 1015 bp in length.
Location/Qualifiers
1. 195217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-50102"
BASE COUNT 61010 a 36076 c 34593 g 59532 t 4006 others
ORIGIN
Query Match 0.7%; Score 21; DB 2: Length 195217;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1146 atacagaaaaaacacacaa 1166
|||||
Db 188931 ATACAGAAAAACAAACAAA 188911
RESULT 42
ATCHRIV83/c
LOCUS ATCHRIV83 197859 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 83.
ACCESSION AL161587
VERSION AL161587.2 GI:7270470
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 16641 to 103350)
Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
Unpublished
2 (bases 98352 to 125759)
Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
3 (bases 120761 to 197859)
Rose,M., Hempel,S., Entlan,K.-D., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
4 (bases 1 to 24256)
Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
Unpublished
5 (bases 1 to 197859)
EU Arabidopsis sequencing,project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4

and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV84 at the 3' end and an
overlap with ATCHRIV82 at the 5' end and an
FEATURES
source
1. 197859
Location/Qualifiers
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
11339..12705
/gene="AT4g35190"
join(11339..11494,11536..11664,11853..11952,12119..12189,
12314..12457,12499..12705)
/gene="AT4g35190"
note="similarity to ctf, Mycobacterium leprae, gb:U15180
contains EST gb:T45691, T22640, T75954"
/codon_start=1
/product="putative protein"
/protein_id="CAB80236.1"
/db_xref="GI:7270471"
/translation="MEIVKSRKRCVFCGSSGKRCESDAATDLAQLVRLCLNLN
ESLENKWTNRNLVYGGSGIGLGLVSAVHEAGHVLGYAQITDLETLTGTETG
EVIAYADMERKAEMARHSDCFIALPGYGTLELLEVLTAQAQGIHDKPVGILVWDG
YNYLITFDKAVDDGFIKPSORHIFVSPAPNAKEVLQLEIMLKINKDKKFPESADLL
FPDIPQAYKPVNDGVYAKSRWEVEKKVQDPQQQQQVFCSSNTSMQTEIAL"
11339..11494
/gene="AT4g35190"
/number=1
11495..11535
/gene="AT4g35190"
/number=1
11536..11664
/gene="AT4g35190"
/number=2
11665..11852
/gene="AT4g35190"
/number=2
11853..11952
/gene="AT4g35190"
/number=3
11953..12118
/gene="AT4g35190"
/number=3
12119..12189
/gene="AT4g35190"
/number=4
12190..12313
/gene="AT4g35190"
/number=4
12314..12457
/gene="AT4g35190"
/number=5
12458..12498
/gene="AT4g35190"
/number=5
12499..12705
/gene="AT4g35190"
/number=6
13757..14518
/gene="AT4g35200"
complement(13757..14518)
/gene="AT4g35200"
complement(13757..14518)
/gene="AT4g35200"
note="similarity to various predicted proteins,
Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB80237.1"
/db_xref="GI:7270472"
/translation="MAVSFVHSNSYPSRQHPQAAVDEQILTRLSSDSASSSSICQ
LSNLDLHDSLEKMRILSVTNALSDQDIEKLIDSLRLIDLCNTAKDAISOMKSGLM
EIOSILRRRPGDLSGEVKYIVSRFLKSLQKVIKSLKVCOSKSDTNAALVFCRAE

ATVMALESLFSEFSGSKACGKSLVSKMNSQNKVTCBAEANEFTRIDSEFQSEKSLQ
MEDVQNLSECTIQLEDEGIESLSKSLIKYRVSIINI"
complement(13157..14518)
/gene="At4g35200"
/number=1
16043..16795
/gene="At4g35210"
16043..16795
/gene="At4g35210"
/number=1
16043..16795
/gene="At4g35210"
/note="similarity to various predicted proteins,
Arabidopsis thaliana
/codon_start=1
/product="putative protein"
/protein_id="CAB80238.1"
/db_xref="GI:7270473"
/translation="MAVSFVRSSTSPSRQHPQAAHYDEQUTLRSSGTAASSSTCOR
LSNDQDLHSLERKMLRSLVFNQALSDQIEKLKLSIKILDLCSIKDLSQMKESIK
EIOISVRRKRGDLSAEVKKYLASKRFLKSEKYLKSLKTSQNKNDALAVGEAEVTV
IALFESLFSFSGSKACGKSLVSKMNSQNKVTCBAEANEFTRIDSEFQSEKSLQMED
VONLEICIQLEDEGIESLSKSLIKYRVSIINI"
16640..24256
/note="position 76883-84499 overlaps to BAC clone F23E12,
EMBL acc:AL022604; for sequence analysis please refer to
this accession"
17252..17657
/gene="At4g35220"
/number=1
17252..18625
/gene="At4g35220"
join(17252..17657,17740..17784,18534..18625)
/gene="At4g35220"
/note="similarity to predicted protein, Arabidopsis
thaliana
contains EST gb:R14002, T46305, T43454, A1996658.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB80239.1"
/db_xref="GI:7270474"
/translation="MAVPLPFLITLISPSLISAGASNAVPSICTAPIDGFTPE
LKRIRRVYGNKRIYDISHRYTPPMSPWDSSEIGRLWMLAASKKNSLANNSEMKIP
THRTGHVDSFGHYDKYDAGFDVDSLQVGLALLVDVPMKNTITMDEQGIQF
VASSSIDSLSPSCQIINCP"
17658..17739
/gene="At4g35220"
/number=1
17740..17784
/gene="At4g35220"
/number=2
17785..18533
/gene="At4g35220"
/number=2
18534..18625
/gene="At4g35220"
/number=3
19941..22657
/gene="At4g35230"
complement(join(19941..20231,20345..20458,20546..20710,
20792..20980,21340..21425,21551..21656,21755..21888,
22078..22213,22304..22657))
/gene="At4g35230"
/note="similarity to protein kinase APK1, Arabidopsis
thaliana, PIR2:S28615
contains EST gb:A1997799.1, F13911"
/codon_start=1
/product="putative protein"
/protein_id="CAB80240.1"

/db_xref="GI:7270475"
/translation="MCCOSLFGSDNPLRGKGVQPPQLSONNHGATTAUNGSGGAS
GVGGGGGGGIPSESEFADIKATNNFSSDNLVSSEGKAPMLVYKGLQNNRWA
VKRFTKMAWPEKPOFAEAMGVGKLRNRLNLIAGCCDGERLLVAFEPNDILAKH
LFHEKNOTITEMAMRLRGVYIAERLDVCSREGRGLHYDIAARVLPEDCDPLRSCG
LMKNSRDGKSYSTLATTPEYLRNGVTPESVYTSRGTYLLDLSKHLPPSHALDM
IRGNKNTLLDMSHLEGFSTEEATVVELLSQCYQYPRRPNKDLVATLAPLQTKS
DVPSYVLMIGIRKOEAPSTPORPLSPGEACSRMDLAIHQIILVMTYRDEGTNELS
FOEWTOQMKDMLDARKRGDOSFREKDKRTIDCSQNSYGIHFCGQFIDVGTWVSP
TVFGRSLCYLLCDOPDPAALRDMAQACVYPMWPLARYMQSVALLAKLNMTVDAADMLN
EAAQLEEKRRGGRGS"
complement(119941..20231)
/gene="At4g35230"
/number=1
complement(20232..20344)
/number=1
complement(20345..20458)
exon
intron
exon
Query Match
Best Local Similarity 100.0%; Score 21; DB 8; Length 197859;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 626 caaatcgctctctccccc 646
Db 91607 CAAATTCGTCCTTCCTCCG 91587
RESULT 43
AC068573/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 198935)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-540B6
2 (bases 1 to 198935)
Anderson,S., Baldwin,D., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodghe,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Leyne,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeckers,R.,
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mienga,V., Morrow,D.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zaitoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 19, 2001 this sequence version replaced gi:10280764.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10276
 Center clone name: 540_E_6
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 41% of reads
 Sequencing vector: Plasmid; n/a; 59% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 19353 bases at least Q40
 Consensus quality: 195810 bases at least Q30
 Consensus quality: 196898 bases at least Q20
 Insert size: 16600; agarose-fp
 Insert size: 197835; sum-of-contigs
 Quality coverage: 10.6 in Q20 bases; agarose-fp
 Quality coverage: 8.9 in Q20.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 4883 4982: contig of 4882 bp in length
 4983 5624: contig of 642 bp in length
 5625 5724: gap of 100 bp
 5725 6734: contig of 1010 bp in length
 6735 6834: gap of 100 bp
 6835 8393: contig of 1559 bp in length
 8394 8493: gap of 100 bp
 8494 11316: contig of 2823 bp in length
 11317 11416: gap of 100 bp
 11417 16171: contig of 4755 bp in length
 16172 16271: gap of 100 bp
 16272 30341: contig of 14070 bp in length
 30342 30441: gap of 100 bp
 30442 49473: contig of 19032 bp in length
 49474 49573: gap of 100 bp
 49574 75583: contig of 26010 bp in length
 75584 75683: gap of 100 bp
 75684 158422: contig of 82739 bp in length
 158423 158522: gap of 100 bp
 158523 180063: contig of 21541 bp in length
 180064 180163: gap of 100 bp
 180164 198935: contig of 18772 bp in length.

FEATURES
 source
 1.198935 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone_lib="RP11-540E6"
 1.4882 /clone_lib="RP11 Human Male BAC"
 misc_feature
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 4983 5624
 misc_feature
 /note="assembly_fragment"
 5725 6734
 misc_feature
 /note="assembly_fragment"
 6835 8393
 misc_feature
 /note="assembly_fragment"
 8494 11316
 misc_feature
 /note="assembly_fragment"
 11417 16171
 misc_feature
 /note="assembly_fragment"
 16272 30341

misc_feature
 /note="assembly_fragment"
 30442 49473
 misc_feature
 /note="assembly_fragment"
 49574 75583
 misc_feature
 /note="assembly_fragment"
 75684 158422
 misc_feature
 /note="assembly_fragment"
 158523 180063
 misc_feature
 /note="assembly_fragment"
 180164 198935
 misc_feature
 /note="assembly_fragment"
 clone_end:77
 vector_side:right"
 BASE COUNT 53896 a 46178 c 45676 g 52083 t 1102 others
 ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 198935;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 caaagctctttcacaagaca 314
 Db 13730 CAAGA GTT TTT CAT CAC GACA 13710
 |||||
 |||||

RESULT 44
 AL593857/c
 LOCUS
 DEFINITION
 PROGRESS *** in unordered pieces.
 ACCESSION
 AL593857
 VERSION
 AL593857.1 GI:14787341
 KEYWORDS
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 206137)
 REFERENCE
 Sims, S.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (11-Jul-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
 requests: clonequest@sanger.ac.uk
 COMMENT
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BM439H2
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: Plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 204045 bases at least Q40
 Consensus quality: 204532 bases at least Q30
 Consensus quality: 204877 bases at least Q20
 Insert size: 205337; sum-of-contigs
 Insert size: 207948; 2.7% error; agarose-fp
 Quality coverage: 10.60x in Q20 bases; sum-of-contigs Quality
 coverage: 10.47x in Q20 bases; agarose-fp

FEATURES
 source
 1.206137 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-439H2"

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

misc_feature      /clone.lib="RPCT-23"
                  1..15760
                  /note="assembly_fragment:02571
                  fragment_chain:1
                  clone_end:SP6
                  vector_side:left"
misc_feature      /note="assembly_fragment:00691
                  fragment_chain:1"
                  19857..46285
                  /note="assembly_fragment:04142
                  fragment_chain:1"
                  46386..146248
                  /note="assembly_fragment:04376
                  fragment_chain:1"
                  146349..150461
                  /note="assembly_fragment:02422
                  fragment_chain:1"
                  150562..161265
                  /note="assembly_fragment:04927
                  fragment_chain:1"
                  161366..165284
                  /note="assembly_fragment:05216
                  fragment_chain:1"
                  165385..186792
                  /note="assembly_fragment:03899
                  fragment_chain:2"
                  186893..206137
                  /note="assembly_fragment:00210
                  fragment_chain:2
                  clone_end:T7
                  vector_side:right"
BASE COUNT      57514 a 47684 c 46179 g 53957 t 803.others
ORIGIN

```

```

Query Match      0.7%; Score 21; DB 2; Length 206137;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1152 aaaaacaacacacacataaa 1172
|||||
Db 63514 AAAAACAAACACAAACATATA 63494

RESULT 45
LOCUS AL513468 211030 bp DNA HTG 21-JUL-2001
DEFINITION Mus musculus chromosome X clone RP23-10416, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL513468
VERSION 1 GI:13872411
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211030)
Tromans, A.
Direct Submission
Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Apr 27, 2001 this sequence version replaced gi:13568261.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseseq@har.mrc.ac.uk
----- Project Information
Center project name: bm10416
----- Summary Statistics
Assembly program: XGAP4; version 4.5

```

```

Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Chemistry: Dye-Primer Big Dye; 1% of reads
Consensus quality: 210707 bases at least Q40
Consensus quality: 210871 bases at least Q30
Consensus quality: 210909 bases at least Q20
Insert size: 210930; sum-of-contigs
Quality coverage: 8.25x in Q20 bases; sum-of-contigs quality
coverage: 8.49x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
SOURCE
  1..211030
  Location/Qualifiers
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /chromosome="X"
    /clone="RP23-10416"
    /clone.lib="RPCT-23"
    1..166819
    /note="assembly_fragment:00499"
    166920..211030
    /note="assembly_fragment:01558"
BASE COUNT      70605 a 40504 c 38146 g 61675 t 100.others
ORIGIN

```

```

Query Match      0.7%; Score 21; DB 2; Length 211030;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1152 aaaaacaacacacacataaa 1172
|||||
Db 135934 AAAAACAAACACAAACATATA 135914

RESULT 46
LOCUS AC092992 221341 bp DNA HTG 09-AUG-2001
DEFINITION Homo sapiens chromosome 3p clone RP11-500K7, WORKING DRAFT
SEQUENCE, 20 unordered pieces.
ACCESSION AC092992
VERSION 1 GI:15137114
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 221341)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alshrocks, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barbata, J.,
Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burck, P., Burrell, C., Burrell, K.L., Byrd, N.C., Carton, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homisi, F., Howard, S., Huber, J., Hulik, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,

```


TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Louisegeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawliny,E., Mcleod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogulu,M., Okunolu,G., Oragunye,N., Ovielo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojurokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
Slisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanek,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 221341)
Worley,K.C.

Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: HDG01
Center clone name: RP11-500K7

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 212184 bases at least Q40
Consensus quality: 217356 bases at least Q30
Consensus quality: 220392 bases at least Q20
Estimated insert size: 220263; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; 4x in Q20 bases; sum-of-coverage estimation
Quality coverage: 4.1x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently
consists of 20 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 22546: contig of 22546 bp in length
22547 22646: gap of unknown length
22647 22647: contig of 25764 bp in length
48411 48510: gap of unknown length
48511 71072: contig of 22562 bp in length
71073 71172: gap of unknown length
71173 91598: contig of 20426 bp in length
91599 91698: gap of unknown length
91699 108834: contig of 17136 bp in length
108835 108934: gap of unknown length
108935 126517: contig of 17583 bp in length
126518 126617: gap of unknown length
126618 139129: contig of 12512 bp in length
139130 139229: gap of unknown length
139230 148601: contig of 9372 bp in length
148602 148701: gap of unknown length
148702 157335: contig of 8634 bp in length
157336 157435: gap of unknown length
157436 166202: contig of 8767 bp in length
166203 166302: gap of unknown length

166303 173060: contig of 6758 bp in length
173061 173160: gap of unknown length
173161 180928: contig of 7768 bp in length
180929 181028: gap of unknown length
181029 188269: contig of 7241 bp in length
188270 188369: gap of unknown length
188370 195914: contig of 7545 bp in length
195915 196015: gap of unknown length
196016 200982: contig of 4878 bp in length
200983 200992: gap of unknown length
200993 205927: contig of 4935 bp in length
205928 206027: gap of unknown length
206028 210855: contig of 4828 bp in length
210856 210955: gap of unknown length
210956 214834: contig of 3879 bp in length
214835 214934: gap of unknown length
214935 217876: contig of 2942 bp in length
217877 221341: contig of 3365 bp in length.
217977 Location/Qualifiers

1. 221341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q"
/clone="RP11-500K7"

BASE COUNT 70894 a 38966 c 37158 g 72403 t 1920 others

ORIGIN

Query Match 0.7%; Score 21; DB 2; Length 221341;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 atcagagaacacacacaa 1166
Db 41072 ATACAGAAAACAAACACAAA 41052

RESULT 47
AC020727
LOCUS Homo sapiens chromosome 3 clone RP11-500K7, WORKING DRAFT SEQUENCE,
DEFINITION 20 unordered pieces.
AC020727
AC020727.5 GI:10048064
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 227194)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 227194)
Waterston,R.H.
Direct Submission
Submitted (08-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 10, 2000 this sequence version replaced gi:7232174.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0500K07
----- Summary Statistics -----
Sequencing vector: M13, 87%
Sequencing vector: plasmid, 13%
Chemistry: Dye-primer ET; 87% of reads
Chemistry: Dye-terminator Big Dye; 13% of reads

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1447: contig of 1447 bp in length
* 1448 1547: gap of unknown length
* 1548 2509: contig of 1062 bp in length
* 2610 2709: gap of unknown length
* 2710 4302 4401: gap of 1592 bp in length
* 4402 7988: contig of 3587 bp in length
* 7989 8088: gap of unknown length
* 8089 10498: contig of 2410 bp in length
* 10499 10598: gap of unknown length
* 10599 13494: contig of 2896 bp in length
* 13495 13594: gap of unknown length
* 13595 17577: contig of 3983 bp in length
* 17578 17677: gap of unknown length
* 17678 20847: contig of 3170 bp in length
* 20848 20947: gap of unknown length
* 20948 24832: contig of 3885 bp in length
* 24833 24933 29910: contig of 4878 bp in length
* 29911 32910: gap of unknown length
* 32911 37423: contig of 7413 bp in length
* 37424 37424 47954: contig of 10531 bp in length
* 47955 48054: gap of unknown length
* 48055 60562: contig of 12508 bp in length
* 60563 76734: contig of 16072 bp in length
* 76735 76835 91907: contig of 15073 bp in length
* 91908 92007: gap of unknown length
* 92008 111826: contig of 19819 bp in length
* 111827 111927 132832: gap of unknown length
* 111927 132832: contig of 21006 bp in length
* 132933 133032: gap of unknown length
* 133033 154508: contig of 21476 bp in length
* 154509 154608: gap of unknown length
* 154609 185758: contig of 31150 bp in length
* 185759 185858: gap of unknown length
* 185859 227194: contig of 41336 bp in length.
*
* Location/Qualifiers
*
* 1..227194
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="3"
*   /clone="RP11-500K7"
*
* 1..1447
*   /note="assembly_name:Contig6"
*
* 1548..2609
*   /note="assembly_name:Contig7"
*
* 2710..4301
*   /note="assembly_name:Contig8"
*
* 4402..7988
*   /note="assembly_name:Contig9"
*
* 8089..10498
*   /note="assembly_name:Contig10"
*
* 10393..13494

```

Query Match	Best Local Similarity	0.7%: Score 21:	DB 2:	Length 227194:
Matches 21:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY 1146 atacagaaaacaacacaaa 1166 	182182 ATACGAAACAAACACAA 182202			
RESULT 48 AE003538	LOCUS Drosophila melanogaster genomic scaffold 142000013386050 section 30 of 54, complete sequence.	INV	06-Oct-2000	
ACCESSION AE003538	VERSION AE003538.2	KEYWORDS GI:10727947		
SOURCE ORGANISM	fruit fly. Drosophila melanogaster Eukaryota; Neoptera; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Endopterygota; Diptera; Brachycera; Muscophora; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 303367)			
REFERENCE AUTHORS	Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., Georgie, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Benois, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Bridon, G.C., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazek, R.G., Champe, M., Pfeiffer, B.D., Man, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews, Pflamkoch, C., Baldwin, D., Ballew, R.W., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brothier, P., Burris, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Dou, P.L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,			

[illegible]

```

mrna
/translation="MSQESGDRSPICVGLGSGOMPKNKETTMTTVLSSPQOKMTM
OMMTLGH"
join(<183391..183408,183461..>183691)
/gene="CG14113"
/product="CT33709"
/db_xref="FLYBASE:FBan0014113"
/db_xref="FLYBASE:FBgn0040814"
/evidence=not_experimental
<183391..>183691
/gene="CG14113"
/map="70A-70A4"
/db_xref="FLYBASE:FBan0014113"
/db_xref="FLYBASE:FBgn0040814"
/evidence=not_experimental
join(183391..183408,183461..183691)
/gene="CG14113"
/note="CG14113 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014113"
/db_xref="FLYBASE:FBgn0040814"
/evidence=not_experimental
/protein_id="AAFA9835.1"
/db_xref="GI:7294492"
/translation="
Query Match      0.7%; Score 21; DB 3; Length 303367;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1666 aaattctgaagcataaca 1686
|||||
Db 162175 AAATTCGAAGCATAACA 162195

RESULT 49
AY027789/c      768 bp      mRNA      PRI      20-JUL-2001
DEFINITION      Homo sapiens CLANC (CLAN1) mRNA, complete cds.
ACCESSION      AY027789
VERSION      AY027789.1 GI:14324116
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 768)
AUTHORS      Damiano,J.S., Stehlik,C., Pio,F., Godzik,A. and Reed,J.C.
TITLE      Cian, a novel human ced-4-like gene
JOURNAL      Genomics. 75 (1-3), 77-83 (2001)
MEDLINE      21365712
PUBMED      11472070
REFERENCE      2 (bases 1 to 768)
AUTHORS      Stehlik,C., Damiano,J.S., Pio,F., Godzik,A. and Reed,J.C.
TITLE      Direct Submission
JOURNAL      Submitted (21-FEB-2001) Program on Apoptosis and Cell Death
Research, The Burnham Institute, 10901 North Torrey Pines Road, La
Jolla, CA 92037, USA
FEATURES
source
Location/Qualifiers
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="p22-p21"
/tissue_type="lung"
1..768
/gene="CLAN1"
277..747
/gene="CLAN1"
/codon_start=1
/product="CLANC"

```

```

BASE COUNT      218 a      157 c      180 g      213 t
ORIGIN
Query Match      0.6%; Score 20; DB 9; Length 768;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2366 agaagatgcataaactag 2385
|||||
Db 755 AGAAGATGCTATAAAGCTAG 736

RESULT 50
AC078439      865 bp      DNA      HTG      02-AUG-2000
LOCUS      Giardia intestinalis clone N55017 strain WB-C6, LOW-PASS SEQUENCE
DEFINITION      SAMPLING.
ACCESSION      AC078439
VERSION      AC078439.1 GI:3654056
KEYWORDS      HTG; PHASEO.
SOURCE      Giardia intestinalis.
ORGANISM      Giardia intestinalis
Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
REFERENCE      1 (bases 1 to 865)
AUTHORS      Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE      Giardia: a model for ancient eukaryotic genome analysis
JOURNAL      Unpublished
2 (bases 1 to 865)
AUTHORS      Nixon,J., Morrison,H.G., McArthur,A.G., Eakin,N.O., Kim,U.,
Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE      Direct Submission
JOURNAL      Submitted (02-AUG-2000) Josephine Bay Paul Center for Comparative
Molecular Biology and Evolution, Marine Biological Laboratory, 7
MBl Street, Woods Hole, MA 02543-1015, USA
COMMENT      * NOTE: This record contains 1 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
Location/Qualifiers
1..865
/organism="Giardia intestinalis"
/strain="WB-C6"
/db_xref="taxon:5741"
/clone="N55017"
BASE COUNT      232 a      218 c      189 g      226 t
ORIGIN
Query Match      0.6%; Score 20; DB 2; Length 865;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 740 gctgctgaagctgcgcaga 759
|||||
Db 800 GCTGCTGAAGCTGCGCAGA 819

RESULT 51

```


FEATURES

Location/Qualifiers

CDS

1..1688
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="II"
complement(397..1410)
/note="ORF YBR141c"
/codon_start=1
/protein_id="CA85099.1"
/db_xref="GI:536440"
/db_xref="SWISS-PROT:P38278"
/translation="MSRSKSTIGTKRKQVSNTRYIKPOKTRIRRHILINKKO
SICKFLCKENLDSDNEKNDKIIILSKGNVRLGKYEGKSSQSFNDAMSQLLRLH
SLIKNEKSKDTSDLAIVYTLGYIMQINKGLGYOIAQSGQLKEGSDTKLL
EKMRSSFPENCGVALEIGSLSSGNRISRCALPRNVYRIDLEHEGVIKODMERPL
PRNENDKFDLISCSLVNPFVKNHRDRCAMCHRMVFLKPGGYFIVIPACVTHSRPC
DKTLQLQLSLISGILMNSHOSNKLTYCYLOLVVSPQSPFSKRIRVNDGPIGLNNG
ITL"

BASE COUNT 468 a 355 c 293 g 572 t
ORIGIN

Query Match 0.6%; Score 20; DB 8; Length 1688;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2955 actaagaattctactcga 2974
|||||
Db 1659 ACTAAGAATTCTACTCGA 1640

RESULT 54
SCYBR142W/c
LOCUS 2833 bp DNA PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome II reading frame ORF YBR142W.
ACCESSION Z36011.131334
VERSION Z36011.1 GI:536441
KEYWORDS
SOURCE
ORGANISM
baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes;
1 (bases 1 to 2833)
Becam,A.M., Herbert,C.J., Naar,F., Slonimski,P.P. and Zagulski,M.
Unpublished
2 (bases 2585 to 2833)
Edlian,K.D., Koeltter,P., Rose,M., Becker,J., Grey,M., Li,Z.,
Niegemann,E., Schenk-Groeninger,R., Servos,J., Weimer,E.,
Wolter,R., Brendel,M., Bauer,J., Braun,H., Dorn,K., Duesterhus,S.,
Gruenbein,R., Hedges,D., Klesau,P., Korol,S., Krems,B., Profft,M.,
Slegers,K., Baur,A., Boles,E., Miosga,T.,
Schaff-Gerstenchaeger,I. and Zimmermann,F.K.
Unpublished
3 (bases 1 to 2833)
MIPS.
Direct Submission
Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
European yeast chromosome II sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried. FRG: E-mail: Mewes@mips.emblnet.org
4 (bases 1 to 2833)
Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Bacle,M.C.,
Barthe,C., Baur,A., Becam,A.M., Bileau,N., Boles,E., Brandt,T.,
Brendel,M., Brueckner,M., Bussereau,F., Christiansen,C.,
Contreras,R., Crouzet,M., Czapluch,C., Demolis,N., Delaveau,T.,
Dolignon,F., Dondy,H., Duesterhus,S., Dubois,E., Dujon,B., El
Bakkoury,M., Entian,K.D., Feuerhahn,M., Fiers,W., Fobo,G.M.,
Fitz,C., Gassenhuber,H., Glansdorff,N., Goffeau,A., Grivell,L.A.,
de Haan,M., Hein,C., Herbert,C.J., Hollenauer,C.P., Holmstrom,K.,
Jacq,C., Jacquet,M., Jauniaux,J.C., Joniaux,J.L., Kallioe,T.,
Klesau,P., Kirchbach,L., Koeltter,P., Korol,S., Liebl,S., Logghe,M.,
Lohan,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Mallet,L.,
Mannhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S.,

Naar,F., Obermaier,B., Perea,J., Pierard,A., Piravandi,E.,
Pohl,F.M., Pohl,T.M., Potier,S., Profft,M., Punelle,B., Ramezani
Rad,M., Rieger,M., Rose,M., Schaff-Gerstenchaeger,I.,
Schierens,B., Schwarzlose,C., Skala,J., Slonimski,P.P.,
Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucke,R.,
Urrestarazu,A., Van der Aart,Q.J., Van Dyck,L., Vassarotti,A.,
Vetter,I., Vierendeels,F., Visser,S., Wagner,G., de Wergifosse,P.,
Wolfe,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kiethe,K.
Complete DNA sequence of yeast chromosome II
EMBO J 13 (24), 5795-5803 (1994)
95112788

FEATURES

Location/Qualifiers

1..2833

CDS

1..2833
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="II"
/gene="MAK5"
279..2600
/gene="MAK5"
279..2600
/gene="MAK5"
/note="ORF YBR142W"
/codon_start=1
/protein_id="CA85100.1"
/db_xref="GI:536442"
/db_xref="SWISS-PROT:P38112"
/translation="MGKKRAPQKQKTVTKNOEIVDESKLNMKPVDPIDTLDGFGFY
GLEIDGVYKVDGKVTFTKQSKYKLNKNSKVDQDQESVENSQSPSEFLER
KNIDDIKEGELSAVSQSDPEDGKNTESKIDPDSQDDEVDYKKNVFKQDN
IDDISPVNDEBNTLAPLAWITIOSLONLFLPRLTELOKSTIPYIMGVYVAKSGY
SGRTLAGVGIPIVEKILNFSQKNKPLSLFTPIRELAHVYTHLKRICKPVAKSGY
SILSTGSLIOKQOURLDNGQIVIAIPGRLELLEKDNLIKRFKSVNLLIDE
ADRLIOGHDFDEFEKILKHLVERKRNENEGSKITWOTLIFSAPESIDLPKLS
ROYVDRFRKNNEEDLNAVIOHLMKIHFNKSPVVIDNPESVSOIKESILCEPPL
RDLYCYEELMPEGTTLIFCNADSVKLVYLNNGIIPAPQIHSSMTOKNRKSLER
FKQSAKQKTIHNSPDVSQSTYLISDVAAKRDIPGQHYIHLPTSDIYIHR
SGRTARAGSEVSAMTCSPOESMGKPLKRTLTATKNSVSTDLNSRTNKKPIKQWT
VPLPIETDILSOLRERSRLAGELADHEIASNSLRKDNMLKRADELGIDVSDDD
ISKNSDTEFLKNNKKNKQKTIKNDKYKAMRATLNELSVPIRDROROKYLTGLVNL
ADNLVKKRGNHSITGHEKTNALFTLKKKKRNN"

BASE COUNT 1002 a 511 c 521 g 799 t
ORIGIN

Query Match 0.6%; Score 20; DB 8; Length 2833;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2955 actaagaattctactcga 2974
|||||
Db 249 ACTAAGAATTCTACTCGA 230

RESULT 55
AF116341/c
LOCUS 3709 bp DNA INV 04-OCT-1999
DEFINITION Drosophila melanogaster poly(A)-binding protein II (Pabp2) gene,
complete cds.
ACCESSION AF116341
VERSION AF116341.1 GI:6007611
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 3709)
Benoit,B., Nemeih,A., Aulner,N., Kuhn,U., Simonelli,M., Wahle,E.
and Bourbon,H.M.
The Drosophila poly(A)-binding protein II is ubiquitous throughout
Drosophila development and has the same function in mRNA
polyadenylation as its bovine homolog in vitro
Nucleic Acids Res. 27 (19), 3771-3778 (1999)
99412419

REFERENCE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

REFERENCE 2 (bases 1 to 3709)
AUTHORS Bourbon,H.-M.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-1998) UMR5540 du CNRS, CNRS, 118 Route de
Narbonne, Toulouse 31062, France

FEATURES
source
1. 3709
/organism="Drosophila melanogaster"
/strain="Oregon R"
/db_xref="taxon:7227"
/chromosome="2"
/map="2R: 448"
join(257..562,761..972,1228..1629,1780..1913,1979..2046,
2472..3226)
/gene="Pabp2"
/product="poly(A)-binding protein II"
257..3226
/gene="Pabp2"
/gene="Pabp2"
join(874..972,1228..1629,1780..1913,1979..2018)
/gene="Pabp2"
/function="RNA-binding protein"
/codon_start=1
/product="poly(A)-binding protein II"
/protein_id="AAE00976.1"
/db_xref="GI:6007612"
/translation="MADEDITNEDDLESEETNGEOTETATEVEEGSMQIDPEL
EATKAVKMEEEAEKIKOMOSEVDKMGSGTGLATVPLSLEKQETDTSVYVGN
VDYGASAELEAHFHCCGTTNNYTLICNKADDPKGFATIEGSKFEVFTALAMNELL
FGRQIKVSKRTNRPGLSTTNFPAFGRFGRARVSRACHSTFGARRAMGYRGA
NYAPY"

BASE COUNT 1146 a 777 c 785 g 1001 t
ORIGIN

Query Match 0.68; Score 20; DB 3; Length 3709;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2534 tgcgaatgcagtgaaatcc 2553
|||||
Db 684 TGCNAATGCAGTGAATTC 665

RESULT 56
SCIRAL/c
LOCUS SCIRAL 12595 bp DNA PLN 22-SEP-1994
DEFINITION S.cerevisiae (S288C) IRA1, YBR1118 and YBR1119 genes.
ACCESSION X78937
VERSION X78937.1 GI:547575
KEYWORDS DEAD box protein; helicase; IRA1 gene; YBR1118 gene; YBR1119 gene.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 12595)
AUTHORS Zagulski,M., Becan,A.M., Grzybowska,E., Lacroite,F., Migdalski,A.,
Slonimski,P.P., Sokolowska,B. and Herbert,C.J.
TITLE The sequence of 12.5 kb from the right arm of chromosome II
predicts a new N-terminal sequence for the IRA1 protein and reveals
two new genes, one of which is a DEAD-box helicase
Yeast 10 (9), 1227-1234 (1994)
JOURNAL 95274325
MEDLINE 2 (bases 1 to 12595)
REFERENCE Herbert,C.J.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1994) C.J. Herbert, Centre de Genetique
Moleculaire, CNRS, 91198 Gif-sur-Yvette, FRANCE
FEATURES
source
1. 12595
/organism="Saccharomyces cerevisiae"
/strain="S288C"
/db_xref="taxon:4932"
/chromosome="II"

gene
complement(1..8331)
/gene="IRA1"
CDS
complement(<1..8331)
/gene="IRA1"
/codon_start=1
/protein_id="CAA55537.1"
/db_xref="GI:547576"
/translation="MNQSDPQKKNPMEYSLTKHLFEDRLILVLPISNLTQYADVE
ADSVNRSCTILNTAITKDLNPITENTIGLIDLYODEETSDNTDITDIAISIVLL
RLSDVFEYEWQNDNDEKIRNDNRKPGKSGSRPMPHRSRPHATRNPLAATMLLCKI
SKLKEFTRLKVLQNMNSHLSGSATISGSIIPDSEOFQKRMNPAYAEIKEDLTIDYI
OREISANHEFTKCVKTVAPLILSHSTKGVNHLIDFGCELDGKRLAATDI
LOHSISYMKRTTFHSLLYAKAKFLFMAKPEYKTYNNLSIDSNVSPSSDNG
GSNSDKTISOLVSLPDDVYSTFSSVSLTNVNDHHTLHSSSSSKTTNTNSPN
STSKTISKOSVNASGNVSPSTGNDSPSPSPASISPLNTNITLGLPLSPITSTL
GQANTSTSTTAATRTDADPTPTMTNNTNNNNNSANLNNIPQRLFSLDDISSFMSR
KSLNLDSSNLFRLMDTSHSNASMTNTNMHAGVNNNSQDSSQSLNEMNIMELYSN
TGSLSHSTAILRFLVLTLLDSEYDEKNSYKRISEPIKINIPKDSNTSMGCSAS
KMPSTRIHLTGKIKTLIOGRRKRNKFLTYLRNLNGQOFVDVSLIDSIRSLFLMT
MPSISQIDSNASVIFSKRFTYNLQNLNLEVTNMSTANTATFISHCYBRNPLFRLRL
QLEFPAISGLQDLSLFLRLQLEKELNHLIDLPKISLYTEGEFVFFHLVSTKLHSDIA
EKTSVLRKLFECIIADILKATPYEDDVNTKLIASLDGHIHLIDQDARTSLNDHVS
PDATSVYTEPEEIIHNSDASLVSLSSPLSINGSGITWTFTWMDIOSILPILSNR
SSASDLSLNLINPLDQONNNANLHARLSGVPTTKRYASPNSERSROSPSSP
QOOSDLPSPSLVSSAGFSSNHSTTATPTPKIKTSKPKTKKIADKOLQPSY
SVYTLSDNDEAKIMANITSIKRNATNFIKRDANTEPKTTDILKPLFVSLDSNQ
RLQVAPARAFETLPYLTAFEDIDNDLDPLVDHLYLCTYAVTLFASPLSLYLENA
KREMLDIIIVKQVRYSVLSNAEKHNLQVALITTEKRLPLLVAGVSGFISLYCSA
RQNTPRLIKISCEFLRSLEFYQKYVGLDOYSIYNIDIFDAMADNFTASGVVALQR
RLRNNTLTFTKSGDSLIDSDMDVIVKKNPFYFSCSKVQOELEVDPRSLAGIASMSGI
LSDMOLEKRSKAPNDEGDSLFSFESNPAYEVHRSKLKLETLKKNPFISKOCOMNIP
NLTRNSRDLTSLHPLPSFLNLSNNIGLKTIDELMSIDLSKSHDSSSVLLQETII
ITRTILKRDDEKIMLFTDLDDAVDKLEIYEKISTISSKYKGTIQNSKFRPEH
SEKNGIASHNHFLLKMKLVLGVGRKLSLNKQYDEENSLRPREMDLQKRDQFLYID
TSEISAKALVLTNHPLEIPSSSEEDNRSTVSFGNHPFILLKGLKESADLQFP
VSLRHKISILNENVIITLNLNANVNSLKFTLPMGSPNDIRIAFLRYVIDVTN
YVNPPEKHEMDKMLAIDDFLKYITKNPLAFGSLACSPADVAGGLNFAFDNRNA
SHIYLETIKOETIKRARSDDLIRKNSCATRLSTYRGRGKKYLIKTLRPVLOCIDV
NKEFTIDMKPQSENSEKMDLPEKRYRLIDATYSSIDDPETELVDICRTTYNAS
VNEPEYAVIAGSVFELRIFGPAVSPDENIITVTHADRPFTLAVLQISLNGR
ENIFKDDILVSKKEELKTCSDKIFNLSLCLKIPNNFVNVNREPTPISEFYSFLAK
FYLNEFTIRKEIINESKLPGSEFLKFTVMNDKILGVAGQSMKEIKNEIPFVEN
REKYSPIEFEMSRVAFKVKYDMKEEEDNAPFAHMAVTLTGIOITVTFNFCYNQNMV
DSLYVKVQIDYARMCSKRYVVIDCTTPGKAGNOKLTTFESLIPBASNCAGCY
YFNVNSFMDQWASSTYENPLYLTTPPCFINSNTDOSLISLGSLSGLSEVLAQVR
VTLHDITLYDKERKRCFVSLKIGNKYEQVLEHIDQLXVTVSNRTFSIKFNNYKIS
NLISVDVNTTGVSSSEFTLSDNEKLYGSKYLEIVKMFYQAQKMEDEGCTGFNS
DISFSTSSAVNASVCNVKVEGEIISHLSLVILVGLFENDDLVKNISYLLAVAGDAF
NIDFGRLHKSPETVVPDDTTFELALIFKAFSSSESTELPPIWKMLOGLENDVPOE
HIPTVYCSLSYVNPNTYEHVYLANDDEGEALSRITISLRTYKRPNTTYAQOIN
FLIADGRLTNVIVETIVSHALDROSDENDKMAVASILTSPTTELACOVKILNMTI
KSFPLAVEASAHWSLTLTISKISVSLFESSPLQWYDEILFVAVSLLDVDPSE
IVSVLYELLMNVCHLTNNESLPERNRKSLMDVOCVAFAROKLNFISFGEQERGLAPN
FAASSFSKFGTIDFTKNIMLMEYXGISEQADWAKKXKYLMDAIFGRHFFSFARA
NMIGIMSKSHSLFLCKRELLVETMKVFAEPVVDQOMIITIAHFYTSKIYEGDPS
SELMKEL"

gene
complement(8728..9741)
/gene="YBR1118"
CDS
complement(8728..9741)
/gene="YBR1118"
/codon_start=1
/protein_id="CAA55538.1"
/db_xref="GI:547577"
/translation="MHSRKSSTGKRRKQSVNVRVINKPQTRRIIRRFHHLINKRQ
SICKFLCKENEDDSSNEKSKDKIIRLSIGANRVLYKDYDGSGOSFNDMEQSLRLH
SLIKNSKSKQTSDLAVAYTTLIGYIMNOIKGLGETYOIASONQOLKRGDGTSKLL
EKWIRSSPENCAGVALALETGSSGRIRISCKLFRNVYVLDILEHGEYKQOPMRPL
PRNENDKPLISCSLVANFVKNRHDTGACHRMVFLAKDGTITFYLPQACQVTHSYIC
DKTLQNLGSLGLIMLNSHOSNKLTYCLYQLOVYVSPSSFSKRIKYNVDGGLNNG
ITL"

250110.
The end of this sequence (40834..40937) overlaps with the start of
sequence 270208.
For a graphical representation of this sequence and its analysis
see: - [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=R01E6)
name=R01E6.

FEATURES

SOURCE

Location/Qualifiers

1..40937

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="X"

/clone="R01E6"

join(999..1127,1467..1589,1647..1756,1803..1905,
1958..2060,2118..2307,2362..2503,2557..2747,3149..3251,
3304..3441,3489..3656,4047..4244)

/gene="R01E6.4"

join(999..1127,1467..1589,1647..1756,1803..1905,
1958..2060,2118..2307,2362..2503,2557..2747,3149..3251,
3304..3441,3489..3656,4047..4244)

/gene="R01E6.4"

/note="contains similarity to Pfam domain: PF00065
(Neurotransmitter-gated ion-channel), Score=484.8,
E-value=2.2e-142, N=1

CDNA EST yk268h8.3 comes from this gene

CDNA EST yk268h8.5 comes from this gene

CDNA EST yk354e9.3 comes from this gene

CDNA EST yk354e9.5 comes from this gene

/codon_start=1

/protein_id="CA92184.1"

/db_xref="GI:3878769"

/translation="MSEFNMALILLVAAGTNAKIRKTAQLESQYEDLLFDYNY
PRPVNSDILTVVGASLIRIIDVENQVLTNNLEMKMNDAKLWPEKYGKLT
LHIPSDFITPDLVLYNNAAGDDITLTDALVTBGNVWOPAIKSPCIDVW
EPYDSCCKEMKFTWYTGKRVYDKQLEQEEVITIKDNDVEFMQOQMDSPFRSA
EMDLSTSEHRSVLYASCGPEKRYVDITVEYGLRRTKLTFTCNILILCPILSTLT
VFYLSDKRTFSPSIIIVTLVFLVLLDMLPSTLVIPMERILITMILVLTYS
VITYNFRSSSAHKMSBPMIYAVLFKLFLKLLMSRPEKGEVTKOPLVNASTLGSN
YATAKANERYRNKAKGNDMLSLRGKPSQAVLNFTDMRMRNNDSTVNRKPYL
CAQNNNAPTVAKRTKROKSKVDVVFEMNLNOVRFJAEHFRNLEGEISDMTVEV
AMVDRFLILFISLVNLTGVTIFLESPLSYDYSKPMNITVPNKFLQGNFYSNNLK"

join(5029..5054,5131..5546,5594..5889)

/gene="R01E6.5"

join(5029..5054,5131..5546,5594..5889)

/gene="R01E6.5"

/note="CDNA EST yk146b3.5 comes from this gene

CDNA EST yk170c5.5 comes from this gene

CDNA EST yk146b3.3 comes from this gene

CDNA EST yk170c5.3 comes from this gene

CDNA EST yk412b8.3 comes from this gene

CDNA EST yk412b8.5 comes from this gene

CDNA EST yk387b5.3 comes from this gene

CDNA EST yk422c1.3 comes from this gene

CDNA EST yk422c1.5 comes from this gene

CDNA EST yk309d12.3 comes from this gene

CDNA EST yk309d12.5 comes from this gene

/codon_start=1

/protein_id="CA92185.1"

/db_xref="GI:3878770"

/translation="MASPISHSHENASGKYGDESOHCKNFHOKAHSHEKAKG
KSDGKAADHDLAKDAEHRKONKEAHOSGSKNYVDASNTNDVKTKTFEFDYR
VOPYHMEQYHTDERHANKYAGDEHNAQOQHRGADGHGYDKAQEHGASHTN
YGNEDAGHSKSYDNEKRGDHTDGYNNKGYDNGGCGHDSSEYQPSYQPKYGGH
QSHYOPARHSSPOGDOYGYEAPH"

join(11014..11029,11105..11235,11321..11420,11472..11686,
11940..12167,12215..12304)

/gene="R01E6.3"

join(11014..11029,11105..11235,11321..11420,11472..11686,
11940..12167,12215..12304)

/gene="R01E6.3"

/note="contains similarity to Pfam domain: PF00194
(Eukaryotic-type carbonic anhydrase), Score=144.1,

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

CDN

gene
CDS
SSLVMTFFSVADGVYSELRRHRMSSRLTKHOSTVDSRENNEQTPGEIHSEVCLHI
PLSPHPIANENVYLTQLRLQOAAQPARSSNNL"
Complement(join(33511..33756,33804..33947,33995..34096,
34158..34242,34362..34369))
/gene="R01E6.2"
34158..34242,34362..34369))
/gene="R01E6.2"
/codon_start=1
/protein_id="CAA92188.1"
/db_xref="GI:3878773"
/db_xref="SPTREMBL:Q21619"

Query Match 0.6%; Score 20; DB 3; Length 40937;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2414 gttgtattcattgacc 2433
Db 22674 GTGTTATTTCATTGACCC 22693

RESULT 59
CBRG39N14/c
LOCUS CBRG39N14 45027 bp DNA INV 04-NOV-2000
DEFINITION Caenorhabditis briggsae cosmid G39N14, complete sequence.
ACCESSION AC084575
VERSION AC084575.1 GI:11095025
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 45027)
Washington University Genome Sequencing Center.
The C. briggsae genome sequencing project
Unpublished
2 (bases 1 to 45027)
Waterston, R.
Direct Submission
Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL
TITLE
AUTHORS
REFERENCE
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

FEATURES
source
1..45027
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="G39N14"
BASE COUNT 14506 a 8254 c 8601 g 13666 t
ORIGIN

Query Match 0.6%; Score 20; DB 3; Length 45027;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 ttcatcagaacatgaatt 42
Db 2861 TTCATCAGACATGATT 2842

RESULT 60

AC006103
LOCUS AC006103 45459 bp DNA HTG 04-DEC-1998
DEFINITION Homo sapiens chromosome 10 clone LA10NC01_124_D_3 map 10q25.1, ***
SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION AC006103
VERSION AC006103.1 GI:3962484
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 45459)
Smith,D.R.
Sequencing of Human Chromosome 10
Unpublished
2 (bases 1 to 45459)
Smith,D.R.
Direct Submission
Submitted (04-DEC-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
JOURNAL
TITLE
AUTHORS
REFERENCE
COMMENT
Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
Technology. Data may contain low quality seq unce and BAC/Cosmid
vector sequences.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 45459: contig of 45459 bp in length.
Location/Qualifiers
1..45459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LA10NC01_124_D_3"
/chromosome="10"
/map="10q25.1"

BASE COUNT 13350 a 9755 c 9636 g 12671 t 47 others
ORIGIN

Query Match 0.6%; Score 20; DB 2; Length 45459;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2914 tattgagaacttaagca 2933
Db 8207 TATTGAGATTCTTAGCAA 8226

RESULT 61
AC005210
LOCUS AC005210 83969 bp DNA PRI 05-NOV-1999
DEFINITION cilt 179_n_3, complete sequence.
ACCESSION AC005210
VERSION AC005210.3 GI:6249673
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 83969)
Smith,D.R.
Sequencing of Human Chromosome 10
Unpublished
2 (bases 1 to 83969)
Smith,D.R.
Direct Submission
Submitted (01-JUL-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL

REFERENCE 3 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

REFERENCE 4 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

REFERENCE 5 (bases 1 to 83969)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA

REMARK COMMENT On Nov 5, 1999 this sequence version replaced gi:4314329.
FEATURES location/Qualifiers
1..83969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_id="CIT987SK-1179N3"
BASE COUNT 24380 a 19036 c 18255 g 22298 t
ORIGIN

Query Match 0.6%; Score 20; DB 9; Length 83969;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2914 tatttgagaatcctaagcaa 2933
|||||
Db 68688 TATTGAGAACTTAAGCAA 68707

RESULT 62
AL159169/c 86155 bp DNA PRI 28-AUG-2000
LOCUS Human DNA sequence from clone RP11-408A13 on chromosome 9, complete sequence.
DEFINITION AL159169
ACCESSION AL159169.14 GI:9944142
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 86155)
AUTHORS Collier,R.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Aug 29, 2000 this sequence version replaced gi:9908927.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human

Chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
RP11-408A13 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-408A13. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true right end of clone RP11-408A13 is at 86155 in this sequence. The true left end of clone RP11-439M12 is at 95 in this sequence. The true right end of clone RP11-120J1 is at 100 in this sequence.

FEATURES location/Qualifiers
source 1..86155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-408A13"
/clone_id="RPCI-11.2"
complement(1..92)
misc_feature 104..499
/note="match: GSS: Em:AQ351225"
misc_feature 113..440
/note="match: GSS: Em:AQ830380"
misc_feature 620..831
/note="match: GSS: Em:AQ720173"
repeat_region 921..1076
/note="106 copies 2 mer tt 55% conserved"
repeat_region 1112..1245
/note="MIR repeat: matches 2. .168 of consensus"
repeat_region 1511..2117
/note="L1M4 repeat: matches 4699. .4836 of consensus"
misc_feature 2745..3053
/note="match: GSS: Em:AQ427218"
repeat_region 3152..3341
/note="AluSg repeat: matches 1. .309 of consensus"
repeat_region 3697..3752
/note="MIR repeat: matches 48. .247 of consensus"
repeat_region 4146..4327
/note="28 copies 2 mer ac 85% conserved"
repeat_region 4708..4847
/note="MER5A repeat: matches 7. .178 of consensus"
repeat_region 4839..5065
/note="MIR repeat: matches 92. .241 of consensus"
repeat_region 5103..5207
/note="MER5A repeat: matches 3. .189 of consensus"
repeat_region 5132..5233
/note="MER5A repeat: matches 65. .166 of consensus"
repeat_region 7190..7282
/note="MER5A repeat: matches 9. .112 of consensus"
repeat_region 7384..7439
/note="L2 repeat: matches 2331. .2421 of consensus"
repeat_region 9060..9293
/note="L2 repeat: matches 2577. .2634 of consensus"
repeat_region 9777..10318
/note="MIR repeat: matches 3. .243 of consensus"
misc_feature 10804..10990
/note="match: GSS: Em:AQ395702"
complement(10404..10831)
misc_feature 14497..14804
/note="match: GSS: Em:AQ030903"
repeat_region 14510..15126
/note="MIR repeat: matches 47. .231 of consensus"
misc_feature 14782..14976
/note="match: GSS: Em:AQ665295"
repeat_region 15854..16584
/note="MIR repeat: matches 66. .256 of consensus"
misc_feature 17012..17077
/note="match: GSS: Em:AQ394111"
repeat_region 17012..17077
/note="L2 repeat: matches 2680. .2745 of consensus"

```

repeat_region 19799..19828
/note="15 copies 2 mer ga 93% conserved"
repeat_region 20263..20342
/note="40 copies 2 mer aa 68% conserved"
repeat_region 23323..23412
/note="12 repeat: matches 2594..2688 of consensus"
repeat_region 24826..24898
/note="MIR repeat: matches 46..119 of consensus"
repeat_region 24991..25058
/note="34 copies 2 mer ga 67% conserved"
repeat_region 25396..25600
/note="MIR repeat: matches 1..217 of consensus"
repeat_region 26063..26184
/note="12 repeat: matches 2572..2710 of consensus"
repeat_region 27687..27775
/note="LIM3 repeat: matches 7659..7739 of consensus"
repeat_region 28585..28788
/note="12 repeat: matches 2558..2749 of consensus"
repeat_region 28834..29034
/note="AluY repeat: matches 1..201 of consensus"
repeat_region 29035..30250
/note="LIM4 repeat: matches 3018..4252 of consensus"
misc_feature 29792..30537
/note="match: GSS: Em:AQ353410"
misc_feature 29795..30640
/note="match: GSS: Em:AQ898185"
misc_feature 30938..31350
/note="match: GSS: Em:AQ627562"
repeat_region 32743..32817
/note="MIR repeat: matches 187..262 of consensus"
repeat_region 33146..33212
/note="12 repeat: matches 2679..2750 of consensus"
repeat_region 35009..35139
/note="LIM5A repeat: matches 6162..6294 of consensus"
repeat_region 35486..35885
/note="MSTB repeat: matches 1..426 of consensus"
repeat_region 35886..37460
/note="MSTB-internal repeat: matches 1..1651 of consensus"
repeat_region 37463..37675
/note="MSTB repeat: matches 203..420 of consensus"
repeat_region 37685..37913
/note="MSTB repeat: matches 1..229 of consensus"
repeat_region 39397..39444
/note="24 copies 2 mer tt 75% conserved"
repeat_region 40066..40259
/note="LIM4 repeat: matches 4385..4593 of consensus"
repeat_region 40534..40833
/note="LIMB3 repeat: matches 5866..6180 of consensus"
repeat_region 41355..41394
/note="20 copies 2 mer aa 82% conserved"
repeat_region 42168..42231
/note="12 repeat: matches 2642..2705 of consensus"
repeat_region 43215..43505
/note="AluY repeat: matches 1..307 of consensus"
repeat_region 43933..43969
/note="12 repeat: matches 2696..2732 of consensus"
repeat_region 43935..43986
/note="MIR repeat: matches 216..260 of consensus"
repeat_region 44879..45241
/note="12 repeat: matches 2264..2709 of consensus"
repeat_region 45313..45472
/note="MIR repeat: matches 4..183 of consensus"
repeat_region 45633..45946
/note="AluY repeat: matches 3..312 of consensus"
repeat_region 46261..46675
/note="HAI1 repeat: matches 1098..1552 of consensus"
repeat_region 46676..46713
/note="19 copies 2 mer aa 81% conserved"
repeat_region 46993..47895
/note="12 repeat: matches 1181..2181 of consensus"
repeat_region 48487..48532
/note="23 copies 2 mer ta 76% conserved"
repeat_region 48581..48646

```

```

repeat_region /note="33 copies 2 mer at 68% conserved"
48824..49187
/note="THE1C repeat: matches 1..371 of consensus"
repeat_region 49538..49616
/note="MERS8B repeat: matches 275..336 of consensus"
repeat_region 49617..49972
/note="LIM2 repeat: matches 5810..6163 of consensus"
misc_feature /note="LIM2 repeat: matches 5810..6163 of consensus"
/note="match: GSS: Em:B88855"
repeat_region 50230..50439
/note="MERS30 repeat: matches 5..230 of consensus"
misc_feature 50459..50934
/note="match: GSS: Em:AQ247716"
repeat_region 51761..52062
/note="MERS33 repeat: matches 3..323 of consensus"
repeat_region 54014..54039
/note="13 copies 2 mer ca 100% conserved"
repeat_region 54903..54981
/note="MIR repeat: matches 51..139 of consensus"

Query Match 0.6% Score 20; DB 9; Length 8615;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1962 tacattccagcagggctgt 1981
|||||
Db 62606 TACATTCCAGCAGGCTGT 62587

RESULT 63
AP000885 86719 bp DNA PRI 16-DEC-1999
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, clone:B680H4, SOD-AML
DEFINITION region, complete sequence.
ACCESSION AP000885
VERSION AP000885.1 GI:6580114
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:B680H4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 86719)
REFERENCE
AUTHORS Fujiyama,A., Yada,T., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Hattori,M., Ishii,K., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 86,719 genomic DNA of 21q22.1
Published Only in Database (1999) in press
2 (bases 1 to 86719)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (15-DEC-1999) to the DDBJ/EMBL/Genbank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(KIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
tel:81-42-778-9923, fax:81-42-778-9924)

FEATURES
Location/Qualifiers
Source 1..86719
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="B680H4"
/map="21q22.1"

BASE COUNT 26401 a 16034 c 16025 g 28259 t
ORIGIN
Query Match 0.6% Score 20; DB 9; Length 86719;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2465 catagtcagctctgtcaa 2484
|||||

```

DB 3109 CATAGTCAGTCTCTGTCAA 3128

RESULT 64
AC005923
LOCUS
DEFINITION Homo sapiens 3p21.3-4 PAC RP4-751E10 (Roswell Park Cancer Institute
AC005923
AC005923
VERSION AC005923.2 GI:4309927
KEYWORDS
SOURCE HTG.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 88326)
Muzny D., Arenson, A.D., Bouck, J., Bunag, C., Chen, J., Chen, Z.,
Culpepper, P., Ding, Y., Dugan, S.P., Durbin, K.J., Forcum, J.,
Ganesh, R.P., Garcia, C., Garcia, D.K., Gorrell, H., Gorrell, L.L.,
He, X., Hernandez, J., Jackson, L.E., Kondejewski, N., Leal, B.,
Lichtarge, O., Liu, W., Logan, O., Lu, J., Martinez, C., Moore, S.,
Moorish, T., Nguyen, N., Oswal, G., Pampell, L.R., Parish, B.J.,
Perez, L.M., Rashid, N.D., Rives, C.M., Scherer, S.E., Shen, H.,
Simon, M.L., Vo, Q.K., Wei, Y., Williamson, A.L., Worley, K., Zhou, X.,
Naylor, S.L., and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 88326)
Worley, K.C.
Direct Submission
Submitted (04-NOV-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 88326)
Worley, K.C.
Direct Submission
Submitted (27-FEB-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 88326)
Worley, K.C.
Direct Submission
Submitted (12-MAR-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 88326)
Worley, K.C.
Direct Submission
Submitted (12-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 1, 1999 this sequence version replaced gi:4033655.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

REFERENCE
AUTHORS
JOURNAL
COMMENT

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features Listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source

QUALSTAT-REPORT

1. 88326
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3p21.3-4"
/clone="RP015-751E10"
complement(1..61)
/rpt_family="Alu"
complement(63..364)
/rpt_family="L1ME"
complement(380..671)
/rpt_family="AluX"
complement(684..827)
/rpt_family="L1ME"
complement(877..966)
/rpt_family="L1ME"
1009..1297
/rpt_family="Aluub"
1299..1327
/rpt_family="AT-rich"
1336..1366
/rpt_family="7SLRNA"
1431..1728
/rpt_family="AluSc"
complement(1772..1914)
/rpt_family="L1ME"
complement(2104..2310)
/rpt_family="AluJo"
complement(2356..2650)
/rpt_family="AluY"
complement(2651..2790)
/rpt_family="L1M4"
complement(2793..2958)
/rpt_family="AluX"
complement(2959..3259)
/rpt_family="AluY"
complement(3262..3402)
/rpt_family="AluX"
complement(3407..3533)
/rpt_family="L1M4"
complement(3554..3785)
/rpt_family="AluSg"
complement(3806..3891)
/rpt_family="L1P"
complement(4023..4181)
/rpt_family="L2"
4182..4486
/rpt_family="AluY"
complement(4567..4610)
/rpt_family="L2"
4833..5049
/rpt_family="AluSg/x"
5733..6014
/rpt_family="AluJo"
6341..6659
/note="Region: Clone qk32c09.x1 Homo sapiens cDNA
AI245778"

```

misc_feature      10007..14335
                  /note="Region: Similar to Human mRNA for KIAA0279 gene
                  DB7469"
repeat_region     complement(12482..12505)
                  /rpt_family="(CA)n"
gene              join(18532..18996,18755..18896,18994..19114,19398..19582,
                  19992..19994,21255..21352,21545..21754,22012..22208,
                  22477..22632,23545..23711,24216..24383,24605..24731,
                  24824..25030,25264..25438,25645..25757,26178..26257,
                  26796..26919,27390..27506,27589..27737,28006..28115,
                  28464..28623,28949..29149,29914..30799,32243..34004)
                  /gene="Homo sapiens mRNA for MEGF2 AB011536"
repeat_region     23873..23986
                  /rpt_family="(CAGA)n"
repeat_region     complement(26062..26140)
                  /rpt_family="(GA)n"
repeat_region     complement(31427..31535)
                  /rpt_family="L2"
repeat_region     31805..32008
                  /rpt_family="L2"
misc_feature      34641..34866
                  /note="Region: Homo sapiens Cpg island DNA genomic MseI
                  fragment 258606"
repeat_region     42837..43001
                  /rpt_family="FAM"
gene              join(43002..43597,43788..43842,44119..44255,44473..44750)
                  /gene="Unigene cluster containing AA057543 and AA411587"
                  44592..44724
                  /gene="Unigene cluster containing AA057543 and AA411587"
STS              45136..45393
                  /db_xref="dbSTS:23502"
                  /standard_name="D3S4222"
repeat_region     45136..45393
                  /rpt_family="AluSx"
repeat_region     45539..45633
                  /rpt_family="AluSx"
repeat_region     45674..45879
                  /rpt_family="MIR"
repeat_region     46117..46224
                  /rpt_family="AluSg/X"
repeat_region     complement(46711..46880)
                  /rpt_family="AluJb"
repeat_region     complement(46901..47180)
                  /rpt_family="AluSx"
repeat_region     complement(47189..47498)
                  /rpt_family="AluY"
repeat_region     complement(47499..47782)
                  /rpt_family="AluSg"
misc_feature      47716..47800
                  /note="Sequence is generated from PCR product only"
                  /function="Low coverage"
misc_feature      47802..48101
                  /function="Gap of approximately 300 bps"
repeat_region     complement(48158..48460)
                  /rpt_family="AluJb"
gene              join(48873..49019,49455..49645)
                  /gene="Unigene cluster containing AA292850 and AA405356"
                  49867..50178
                  /rpt_family="AluJb"
repeat_region     complement(50948..51140)

Query Match      0.6%; Score 20; DB 9; Length 88326;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1961 ctacattccagcaggcgtc 1980
      |||||||
Db 72992 CTACATTCACGACGCGCTG 73011

RESULT 65
AC084447 88839 bp DNA INV 04-NOV-2000
LOCUS AC084447/c
DEFINITION Caenorhabditis briggsae cosmid CB019G12, complete sequence.

```

```

ACCESSION AC084447
VERSION AC084447.1 GI:11094897
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae
          Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 88839)
          Washington University Genome Sequencing Center.
          The C. briggsae Genome Sequencing Project
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 88839)
TITLE Waterston, R.
JOURNAL Direct Submission
COMMENT Submitted (04-NOV-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          Submitted by:
          Genome Sequencing Center
          Department of Genetics, Washington University,
          St. Louis, MO 63110, USA
          e-mail: jsplethwatson.wustl.edu

FEATURES
Source
Location/Qualifiers
1..88839
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="CB019G12"
9330..9401
/note="codon recognized: CCA"
tRNA /product="tRNA-Pro"
9814..9885
/note="codon recognized: CCU"
tRNA /product="tRNA-Pro"
30927..30998
/note="codon recognized: GGA"
tRNA /product="tRNA-Gly"
28080 a 16056 c 16645 g 28058 t
BASE COUNT
ORIGIN

Query Match      0.6%; Score 20; DB 3; Length 88839;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2749 aggtcccaactcgtcaag 2768
      |||||||
Db 67631 AGGTCCCACTCGTCAAG 67612

RESULT 66
AL391278/c 91733 bp DNA HTG 13-JUN-2001
LOCUS AL391278/c
DEFINITION Homo sapiens chromosome 1 clone RP5-885P2, *** SEQUENCING IN
          PROGRESS ***; 29 unordered pieces.
ACCESSION AL391278
VERSION AL391278.10 GI:12331116
KEYWORDS HTG; HTGS-PHASE1; HTGS-CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 91733)
          McIay, K.
JOURNAL Direct Submission
COMMENT Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          requests: clonerequest@sanger.ac.uk

```

----- Genome Center
On Jan 22, 2001 this sequence

Center: Sanger Centre
Center code: SC

Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

```
----- Project Information
Center project name: dJ885P2
```

```
----- Summary Statistics
Assembly program: XGAP4; version 4.5
```

Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 14% of reads Chemistry:

Dye-terminator Big Dye; 85% of reads
Consensus quality: 76799 bases at least Q40

Consensus quality: 81861 bases at least Q30
Consensus quality: 85494 bases at least Q20

Insert size: 88933; sum-of-contigs
Insert size: 169449; 2.68 error; agarose-fp

Quality coverage: 2.56x in Q20 bases; sum-of-contigs Quality coverage: 1.81x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will then be preserved.

1	2438	contig	of	2438	bp	in	length
2	2439	gap	of	100	bp		
3	2539	4657	contig	of	2119	bp	in
4	4658	4757	gap	of	100	bp	in
5	4758	7026	contig	of	2269	bp	in
6	7027	7126	gap	of	100	bp	in
7	7127	9886	contig	of	2760	bp	in
8	9887	9986	gap	of	100	bp	in
9	9987	12786	contig	of	2800	bp	in
10	12787	12886	gap	of	100	bp	in
11	12887	15825	contig	of	2939	bp	in
12	15826	15925	gap	of	100	bp	in
13	15926	17982	contig	of	2057	bp	in
14	17982	18082	gap	of	100	bp	in
15	18083	24394	contig	of	6312	bp	in
16	24395	24494	gap	of	100	bp	in
17	24495	27186	contig	of	2692	bp	in
18	27187	27286	gap	of	100	bp	in
19	27287	30217	contig	of	2885	bp	in
20	30217	30271	gap	of	100	bp	in
21	30272	33949	contig	of	2678	bp	in
22	33950	33049	gap	of	100	bp	in
23	33050	36523	contig	of	3474	bp	in
24	36524	36623	gap	of	100	bp	in
25	36624	38651	contig	of	2028	bp	in
26	38652	38751	gap	of	100	bp	in
27	38752	41483	contig	of	2732	bp	in
28	41484	41583	gap	of	100	bp	in
29	41584	44333	contig	of	2750	bp	in
30	44334	44433	gap	of	100	bp	in
31	44434	47116	contig	of	2683	bp	in
32	47117	47216	gap	of	100	bp	in
33	47217	48339	contig	of	2123	bp	in
34	48340	49439	gap	of	100	bp	in
35	49440	54013	contig	of	4574	bp	in
36	54014	54113	gap	of	100	bp	in
37	54114	57401	contig	of	3288	bp	in
38	57402	57501	gap	of	100	bp	in
39	57502	60611	contig	of	3110	bp	in
40	60612	60711	gap	of	100	bp	in
41	60712	63058	contig	of	2347	bp	in
42	63059	63158	gap	of	100	bp	in
43	63159	65257	contig	of	2099	bp	in
44	65258	65357	gap	of	100	bp	in

FEATURES	source
*	65358 67754: contig of 2397 bp in length
*	67755 67854: gap of 100 bp
*	67855 72100: contig of 4246 bp in length
*	72101 72200: gap of 100 bp
*	72201 75487: contig of 3287 bp in length
*	75488 75587: gap of 100 bp
*	75588 80699: contig of 5112 bp in length
*	80700 80799: gap of 100 bp
*	80800 84813: contig of 4014 bp in length
*	84814 84913: gap of 100 bp
*	84914 87462: contig of 2549 bp in length
*	87463 87562: gap of 100 bp
*	87563 91733: contig of 4171 bp in length
	Location/Qualifiers
	1..91733

misc_feature	1. 2438 /note=assembly-fragment:00813 fragment_chain:1"
misc_feature	2539. 4657 /note=assembly-fragment:00179 fragment_chain:1"
misc_feature	4758. 7026 /note=assembly-fragment:00827 fragment_chain:2"
misc_feature	7127. 9886 /note=assembly-fragment:00902 fragment_chain:2"
misc_feature	9987. 12786 /note=assembly-fragment:00136 12887. 15825
misc_feature	/note=assembly-fragment:00196 15926. 17982
misc_feature	/note=assembly-fragment:00205 18083. 24394
misc_feature	/note=assembly-fragment:00289 24495. 27186
misc_feature	/note=assembly-fragment:00290 27287. 30171
misc_feature	/note=assembly-fragment:00329 30212. 32949
misc_feature	/note=assembly-fragment:00334 33050. 36523
misc_feature	/note=assembly-fragment:00337 36624. 38651
misc_feature	/note=assembly-fragment:00435 38752. 41483
misc_feature	/note=assembly-fragment:00437 41584. 44333
misc_feature	/note=assembly-fragment:00476 44434. 47116
misc_feature	/note=assembly-fragment:00505 47217. 49339
misc_feature	/note=assembly-fragment:00589 49440. 54013
misc_feature	/note=assembly-fragment:00596 54114. 57401
misc_feature	/note=assembly-fragment:00665 57502. 60611
misc_feature	/note=assembly-fragment:00677 60712. 63058
misc_feature	/note=assembly-fragment:00681 63159. 65257
misc_feature	/note=assembly-fragment:00758 65358. 67754
misc_feature	/note=assembly-fragment:00805 67855. 72100
misc_feature	/note=assembly-fragment:00835 72201. 75487

misc_feature /note="assembly_fragment:00879"
75588.80699
misc_feature /note="assembly_fragment:00912"
80800.84813
misc_feature /note="assembly_fragment:00938"
84914.87462
misc_feature /note="assembly_fragment:00953"
87563.91733
/note="assembly_fragment:00807
clone_end:SP6
vector_side:right"
BASE COUNT 25532 a 18159 c 18530 g 26693 t 2819 others
ORIGIN

Query Match 0.6%; Score 20; DB 2: Length 91733;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2596 attatcagaattaccgt 2615
Db 88326 ATTATCGAATAATTACTCG 88307
|||||

RESULT 67
AC073294
LOCUS AC073294 93409 bp DNA HTG 26-MAR-2001
DEFINITION Mus musculus clone CT7-32119 strain 129 SV, WORKING DRAFT SEQUENCE,
1 ordered pieces.
ACCESSION AC073294
VERSION AC073294.1 GI:8493569
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 93409)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
2 (bases 1 to 93409)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

FEATURES
source
1. 93409
/organism="Mus musculus"
/strain="129 SV"
/db_xref="taxon:10090"
/clone="CT7-32119"
/clone_lib="CltDCJ7 mouse BAC library"
BASE COUNT 25162 a 21194 c 20677 g 25391 t 985 others
ORIGIN

Query Match 0.6%; Score 20; DB 2: Length 93409;
Best Local Similarity 100.0%; Pred. No. 76;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2172 aacattatctctcatggt 2191
Db 76311 AACATTATCTCTCATGCT 76330
|||||

RESULT 68
AC083819/c
LOCUS AC083819 102165 bp DNA HTG 06-NOV-2000
DEFINITION Mus musculus chromosome 1 clone RP23-285F20, *** SEQUENCING IN
PROGRESS ***; 7 unordered pieces.
ACCESSION AC083819
VERSION AC083819.4 GI:10803522
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 102165)
AUTHORS Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwunonu, G., Carllock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Fansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Louissegh, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mel, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Wellington, S.,
Williamson, A., Wrenstford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,
Worley, K., and Gibbs, R.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 102165)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----Project Information
Center project name: MARI
Center clone name: RP23-285F20
-----Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 95179 bases at least Q40
Consensus quality: 98007 bases at least Q30
Consensus quality: 99497 bases at least Q20
Estimated insert size: 100390; sum-of-contris estimation
Quality coverage: 0x in Q20 bases; agrose-fp estimation
Quality coverage: 2.9x in Q20 bases; sum-of-contris estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

* 1 24502: contig of 24502 bp in length
* 24503 24602: gap of unknown length
* 24603 40809: contig of 16207 bp in length
* 40810 40909: gap of unknown length
* 40910 63389: contig of 22480 bp in length
* 63390 63489: gap of unknown length
* 63490 78501: contig of 15012 bp in length
* 78502 89787: gap of unknown length
* 89788 89887: gap of unknown length
* 89888 96681: contig of 6794 bp in length
* 96682 96781: gap of unknown length
* 96782 102165: contig of 5384 bp in length.

```

FEATURES

```

source
  1..102165
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /chromosome="1"
    /clone="RP23-285F20"

```

```

BASE COUNT 30175 a 21419 c 20896 g 29070 t 605 others
ORIGIN

```

```

Query Match 0.6%; Score 20; DB 2; Length 102165;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 2055 aagcaagatcacatct 2074
|||||
Db 51969 AAGCAAGATCACATCT 51950

```

```

RESULT 69
AL357352 106601 bp DNA PRI 20-NOV-2000
LOCUS Human DNA sequence from clone Rpl1-799G19 on chromosome 6, complete
sequence.
ACCESSION AL357352
VERSION AL357352.11 GI:11321993
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 106601)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

```

COMMENT

On Nov 23, 2000 this sequence version replaced gi:11225761. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the validation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>
 Rpl1-799G19 is from the library RPl1-11.3 constructed at the

Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pbac3.6>

IMPORTANT: This sequence is not the entire insert of clone Rpl1-799G19 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone Rpl1-799G19 is at 1 in this sequence. The true left end of clone Rpl1-532M4 is at 106502 in this sequence.

FEATURES

```

source
  1..106601
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /clone="Rpl1-799G19"
    /clone_lib="RPl1-11.3"
  8..506
    /note="match: GSS: Em:A0521292"
  961..2386
    /note="TIGER2 repeat: matches 1..1483 of consensus"
  2387..2755
    /note="TIGER2 repeat: matches 1..371 of consensus"
  2756..4009
    /note="TIGER2 repeat: matches 1483..2718 of consensus"
  4387..4995
    /note="LIP3 repeat: matches 768..1393 of consensus"
  4995..6352
    /note="LIP12 repeat: matches 4779..6163 of consensus"
  9015..9156
    /note="L2 repeat: matches 2608..2750 of consensus"
  9690..9802
    /note="MIR repeat: matches 56..172 of consensus"
  10378..10458
    /note="MIR repeat: matches 71..153 of consensus"
  10787..10832
    /note="23 copies 2 mer at 78% conserved"
  11644..11807
    /note="L2 repeat: matches 2581..2749 of consensus"
  11817..11984
    /note="MIR repeat: matches 5..186 of consensus"
  12435..12545
    /note="MIR repeat: matches 8..128 of consensus"
  13942..13995
    /note="18 copies 3 mer tta 88% conserved"
  13954..13998
    /note="3 copies 15 mer 93% conserved"
  13996..14278
    /note="Alu repeat: matches 1..286 of consensus"
  14302..14364
    /note="L2 repeat: matches 2634..2701 of consensus"
  14716..14753
    /note="19 copies 2 mer ac 81% conserved"
  15675..15806
    /note="FLAN repeat: matches 1..133 of consensus"
  17568..17732
    /note="L2 repeat: matches 1683..1850 of consensus"
  17850..18255
    /note="match: GSS: Em:A0790581"
  19598..19685
    /note="MLT1J repeat: matches 221..337 of consensus"
  20418..20555
    /note="MER5A repeat: matches 47..182 of consensus"
  29307..29439
    /note="MIR repeat: matches 87..245 of consensus"
  29675..30120
    /note="MLTIF repeat: matches 91..529 of consensus"
  30350..30644
    /note="AluSg repeat: matches 1..313 of consensus"
  30687..30860
    /note="L2 repeat: matches 1728..1906 of consensus"
  33081..33579
    /note="match: GSS: Em:A0783759"
  33928..34076

```

```

/note="MIR repeat: matches 4. .165 of consensus"
34988. .35085
/note="LIP3 repeat: matches 6048. .6146 of consensus"
35092. .35173
/note="LIMC2 repeat: matches 6245. .6328 of consensus"
35174. .35517
/note="MLT1A1 repeat: matches 1. .354 of consensus"
35525. .36385
/note="LIMC1 repeat: matches 5357. .6250 of consensus"
36388. .37713
/note="LIM4 repeat: matches 3807. .5147 of consensus"
40703. .40763
/note="L2 repeat: matches 2676. .2734 of consensus"
40918. .41362
/note="match: GSS: Em:AQ756546"
42175. .42385
/note="match: GSS: Em:AQ113333"
/note="match(42487. .42933)
/note="match: GSS: Em:AQ188019"
complement(42506. .42945)
/note="match: GSS: Em:AQ408008"
complement(43648. .44149)
/note="match: GSS: Em:AQ563645"
complement(43831. .44224)
/note="match: GSS: Em:A2335395"
44050. .44510
/note="match: GSS: Em:AQ117470"
44201. .44518
/note="match: GSS: Em:AQ710247"
45984. .46189
/note="LIM4C repeat: matches 1430. .1645 of consensus"
46033. .46435
/note="match: GSS: Em:AQ622979"
46036. .46436
/note="match: GSS: Em:AQ622990"
46385. .46820
/note="LIMC1 repeat: matches 4942. .5389 of consensus"
46829. .47635
/note="LIMC1 repeat: matches 3969. .4776 of consensus"
47842. .47889
/note="24 copies 2 mer ta 97% conserved"
47844. .47897
/note="3 copies 18 mer 85% conserved"
47890. .48044
/note="LIP repeat: matches 3526. .3683 of consensus"
48082. .48365
/note="LIM4 repeat: matches 3095. .3390 of consensus"
49813. .50122
/note="AluSg repeat: matches 1. .295 of consensus"
50653. .50930
/note="L2 repeat: matches 2008. .2286 of consensus"
51027. .51334
/note="AluY repeat: matches 1. .308 of consensus"
51368. .51496
/note="LIP1A10 repeat: matches 6029. .6156 of consensus"
51510. .51947
/note="L2 repeat: matches 1292. .1752 of consensus"
52009. .52098
/note="45 copies 2 mer ga 84% conserved"
52028. .52099
/note="4 copies 18 mer 91% conserved"
52282. .52518
/note="LIP27 repeat: matches 5559. .5793 of consensus"
52519. .52860
/note="LIP47 repeat: matches 5789. .6143 of consensus"
53493. .54159
/note="match: GSS: Em:B93480"
53511. .53921
/note="match: GSS: Em:AQ011862"
55163. .55556
/note="MERS7B repeat: matches 1. .403 of consensus"
55697. .56012
/note="AluSc repeat: matches 1. .305 of consensus"

```

```

misc_feature 56584. .57061
/note="match: GSS: Em:AQ887852"
56588. .56963
misc_feature /note="match: GSS: Em:AQ824739"
56874. .56931
repeat_region /note="29 copies 2 mer gt 70% conserved"
57865. .58157
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
58991. .59162
repeat_region /note="MERSA repeat: matches 6. .189 of consensus"
59736. .59927
repeat_region /note="MERSB repeat: matches 1. .177 of consensus"
59928. .60263
/note="L2 repeat: matches 144. .499 of consensus"
60293. .60745

```

```

Query Match 0.6%; Score 20; DB 9; Length 106601;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 306 catcagacatcagaagaga 325
Db 58795 CATCAGACATCAGAGAGA 58814

```

```

RESULT 70
AL139036/c AL139036 109810 bp DNA PRI 26-MAY-2001
LOCUS Human DNA sequence from clone RP11-285H24 on chromosome
DEFINITION 13q22.1-22.3, complete sequence.
ACCESSION AL139036
VERSION AL139036.12 GI:14270405

```

```

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 109810) Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Leonagamort, D.

```

```

TITLE Direct Submission
JOURNAL Submitted (26-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

```

```

COMMENT requests: clone request@sanger.ac.uk
On May 31, 2001 this sequence version replaced gi:14269884.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13> RP11-285H24 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-285H24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-810G21 is at 109711 in this sequence. The true right end of clone RP11-157H4 is at 100 in this sequence.

```
FEATURES
source          Location/Qualifiers
                1..109810
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="13"
                /map="q22.1-22.3"
                /clone="RP11-285H24"
                /clone_lib="RPC1-11.1"
                complement(1..98)
misc_feature    /note="match: GSS: Em:AQ374965"
                205..514
                /note="match: GSS: Em:AQ092767"
                1091..1201
                /note="MLT1H repeat: matches 81..189 of consensus"
                1395..1522
                /note="MLT1H repeat: matches 409..510 of consensus"
                2642..3032
                /note="match: GSS: Em:AQ477479"
                2718..3197
                /note="MLT1D repeat: matches 1..502 of consensus"
                3752..3940
                /note="7 copies 27 mer 66% conserved"
                3773..4126
                /note="59 copies 6 mer 99999a 54% conserved"
                3817..4088
                /note="16 copies 17 mer 59% conserved"
                3818..4076
                /note="7 copies 37 mer 63% conserved"
                5037..5248
                /note="MIR repeat: matches 12..248 of consensus"
                complement(6151..6886)
                /note="match: GSS: Em:AK519175"
                complement(6237..6883)
                /note="match: GSS: Em:AQ283095"
                complement(6380..6889)
                /note="match: GSS: Em:AQ581403"
                6623..6699
                /note="HAL1 repeat: matches 669..746 of consensus"
                7701..7870
                /note="MIR repeat: matches 60..220 of consensus"
                8648..8683
                /note="18 copies 2 mer tt 80% conserved"
                complement(9052..9485)
                /note="match: GSS: Em:AQ620944"
                9149..9432
                /note="Alusx repeat: matches 1..284 of consensus"
                9905..10103
                /note="MER63A repeat: matches 1..210 of consensus"
                10116..10217
                /note="L1ME3A repeat: matches 6016..6116 of consensus"
                11190..11776
                /note="match: GSS: Em:AZ3737306"
                11196..11815
                /note="match: GSS: Em:AZ374726"
                13390..13554
                /note="MIR repeat: matches 12..168 of consensus"
                15651..15963
                /note="AluY repeat: matches 1..306 of consensus"
                18346..18991
                /note="match: GSS: Em:AZ516011"
                18347..18878
                /note="match: GSS: Em:AQ384942"
                18568..18864
                /note="Aluub repeat: matches 3..306 of consensus"
                complement(18803..19250)
                /note="match: GSS: Em:AQ801630"
                20484..20699
                /note="MIR repeat: matches 11..252 of consensus"
                23381..23476
                /note="L1M4 repeat: matches 4013..4108 of consensus"
```

```
repeat_region  23477..23773
                /note="AluY repeat: matches 1..299 of consensus"
repeat_region  23774..24043
                /note="L1M4 repeat: matches 4107..4379 of consensus"
                24092..24357
                /note="L1ME3 repeat: matches 5861..6131 of consensus"
                complement(24264..24689)
                /note="match: GSS: Em:AQ451631"
                24793..24874
                /note="MIR repeat: matches 67..145 of consensus"
                25008..25518
                /note="match: GSS: Em:AQ517147"
                25611..25698
                /note="MIR repeat: matches 2..91 of consensus"
                26142..26439
                /note="Alusx repeat: matches 3..307 of consensus"
                27548..27704
                /note="MER33 repeat: matches 22..166 of consensus"
                27705..28031
                /note="MER2 repeat: matches 4..345 of consensus"
                28032..28156
                /note="MER33 repeat: matches 166..324 of consensus"
                30049..30177
                /note="FLM_A repeat: matches 1..133 of consensus"
                31105..31202
                /note="MIR repeat: matches 12..111 of consensus"
                31520..31675
                /note="MIR repeat: matches 100..261 of consensus"
                32359..32560
                /note="L1M4 repeat: matches 5318..5518 of consensus"
                32561..32861
                /note="Aluuo repeat: matches 1..296 of consensus"
                32862..33218
                /note="L1M4 repeat: matches 5518..5905 of consensus"
                33290..33388
                /note="L1ME3A repeat: matches 6029..6129 of consensus"
                33458..33489
                /note="16 copies 2 mer aa 100% conserved"
                33902..33951
                /note="25 copies 2 mer at 74% conserved"
                34308..34780
                /note="L1M48 repeat: matches 5785..6288 of consensus"
                35285..35684
                /note="match: GSS: Em:AQ064667"
                36808..37104
                /note="AluY repeat: matches 1..296 of consensus"
                37255..37308
                /note="2 copies 27 mer 94% conserved"
                37728..37913
                /note="MER63A repeat: matches 1..210 of consensus"
                38654..38727
                /note="2 copies 37 mer 100% conserved"
                39086..39226
                /note="MER3B repeat: matches 29..158 of consensus"
                39258..39424
                /note="MER3B repeat: matches 2..172 of consensus"
                39564..39981
                /note="MLT1C repeat: matches 17..466 of consensus"
                44189..44535
                /note="L1MC/D repeat: matches 5436..5813 of consensus"
                45138..45321
                /note="MIR repeat: matches 81..260 of consensus"
                49524..49727
                /note="MIR repeat: matches 24..252 of consensus"
                49821..50326
                /note="match: GSS: Em:AQ169074"
                49836..50027
                /note="L1MC/D repeat: matches 5611..5813 of consensus"
                50250..50567
                /note="Alusx repeat: matches 1..312 of consensus"
                51288..51497
                /note="MER2 repeat: matches 1..209 of consensus"
                52175..52231
                repeat_region
```

repeat_region /note="MIR repeat: matches 15..74 of consensus"
53123..53421
/note="AlusA1 repeat: matches 1..299 of consensus"
repeat_region 54142..54423
/note="L1MD2 repeat: matches 6011..6278 of consensus"
misc_feature 54454..55060
/note="match: GSS: Em:AQ195363"
repeat_region 54535..54652

Query Match 0.6%; Score 20; DB 9; Length 109810;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2826 ttgcgaagaccctcga 2845
|||||
Db 62720 TTTCGAAGAACCTCTGAA 62701

RESULT 71
LOCUS AC010902 114169 bp DNA PRI 25-MAR-2001
DEFINITION Homo sapiens clone Rpl1-549J7, complete sequence.
ACCESSION AC010902
VERSION AC010902.4 GI:10440867
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 114169)
TITLE The sequence of Homo sapiens clone
JOURNAL 2 (bases 1 to 114169)
AUTHORS Waterston,R.H.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 114169)
AUTHORS Waterston,R.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 114169)
AUTHORS Waterston,R.H.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (25-MAR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT
On Sep 30, 2000 this sequence version replaced gi:7631065.
Center Project name: H_NH0549J07.

FEATURES
source 1..114169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rpl1-549J7"

BASE COUNT 34392 a 22524 c 23021 g 34232 t
ORIGIN

Query Match 0.6%; Score 20; DB 9; Length 114169;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 aaagcttttcacagaca 314
|||||
Db 74191 AAAGCTTTTTCATCAGACA 74172

RESULT 72
LOCUS AC022124/c
AC022124 120733 bp DNA PRI 21-JUL-2001

DEFINITION Homo sapiens chromosome 5 clone CTD-2075619, complete sequence.
ACCESSION AC022124
VERSION AC022124.6 GI:14993687
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 120733)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE
AUTHORS 2 (bases 1 to 120733)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 120733)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jul 21, 2001 this sequence version replaced gi:14277276.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.

FEATURES
source 1..120733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2075619"

BASE COUNT 39434 a 20042 c 22104 g 39153 t
ORIGIN

Query Match 0.6%; Score 20; DB 9; Length 120733;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 400 aagatattgacattttt 419
|||||
Db 61200 AAGATATTGACATTATTTT 61181

RESULT 73
LOCUS AC008390 121931 bp DNA PRI 20-DEC-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-235M3, complete sequence.
ACCESSION AC008390
VERSION AC008390.7 GI:11908269
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 121931)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE
AUTHORS 2 (bases 1 to 121931)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

REFERENCE      3 (bases 1 to 121931)
AUTHORS        DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE          Direct Submission
JOURNAL        Submitted (20-DEC-2000) DOE Joint Genome Institute, 2800 Mitchell
                Drive, Walnut Creek, CA 94598, USA
                On Dec 20, 2000 this sequence version replaced gi:7711258.
COMMENT        Draft Sequence produced by DOE Joint Genome Institute
                www.jgi.doe.gov
                Finishing Completed at Stanford Human Genome Center
                www.shgc.stanford.edu
                Quality: Phrap Quality >=40 99.8% of Sequence;
                Estimated Total Number of Errors is 0.2.

FEATURES
  Source
    1..121931
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="CTC-235M3"

BASE COUNT     39800 a 20269 c 22375 g 39487 t

ORIGIN
Query Match    0.6%; Score 20; DB 9; Length 121931.
Best Local Similarity 100.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 400 aagatattgacattattt 419
    |||||||
Db 62415 AAGATTATGACATATTTT 62396

RESULT 74
AC010072/2c    124347 bp    DNA    PRI    19-NOV-1999
DEFINITION     Homo sapiens chromosome 14q31 clone CTD-217314 containing TSHR
                gene, partial cds; and unknown gene, complete sequence.
ACCESSION      AC010072
VERSION        AC010072.5 GI:6453843
KEYWORDS       HTG.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 124347)
AUTHORS        Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
                Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G.,
                James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shaffer, T.
                and Hood, L.
                Sequencing of human chromosome 14
                Unpublished
                2 (bases 1 to 124347)
                Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
                Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S.,
                Madan, A., Ratcliffe, A., Shaffer, T. and Hood, L.
                Direct Submission
                Submitted (11-SEP-1999) Multimegabase Sequencing Center, University
                of Washington, PO BOX 357730, Seattle, WA 98195, USA
                3 (bases 1 to 124347)
                Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradaran, L., Birditt, B.,
                Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G.,
                James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shaffer, T.
                and Hood, L.
                Direct Submission
                Submitted (19-NOV-1999) Multimegabase Sequencing Center, University
                of Washington, PO BOX 357730, Seattle, WA 98195, USA
                On Nov 19, 1999 this sequence version replaced gi:6114900.

TITLE          JOURNAL
COMMENT        Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chrtona.mbt.washington.edu/msg_www
                Contact: leerowen@u.washington.edu
                ----- Summary Statistics -----
                Sequencing vector: pUC18; L08752

FEATURES
  source
    1..124347
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="14"
      /map="14q31"
      /clone="CTD-217314"
      /clone_1ib="Caltech human BAC library D"
      /note="this clone overlaps RP11-114N19, Accession
      AC007262"
      complement(join(42575..42673,43714..43786,47934..48025,
      52809..52927,53951..54077,62318..62371))
      /note="Intron-exon boundaries defined in relation to EST
      A154064, the closest BLASTX similarity is to a
      drosophila potassium channel protein"
      /codon_start=1
      /product="unknown"
      /protein_id="AAF09033.1"
      /db_xref="GI:6453845"
      /translation="MIGRYREYNGCAATIEIKESLEOSTIDLRSGRLNSGGRST
      SVTSLSASDLGGTGSSELHHPPTSPKDYGDPOGIKMRSRGTGVRFVQETDMDQLH
      GFHSLRDLSSEQIRLGDENRSLRSRSDAETKRALELTETKLNFAQKQEVFFVNS
      LGKVKIQCLSAEVDNVPINWOCGI"
      95810..95860
      /note="low quality data"
      103634..>103846
      /note="this is the 5' end of TSHR, found in M73745"
      /product="thyroid stimulating hormone receptor"
      103677..>103846
      /note="thyroid stimulating hormone receptor"
      /codon_start=1
      /product="TSHR"
      /protein_id="AAF09032.1"
      /db_xref="GI:6453844"
      /translation="MRPADLQVLVLLDLPRLDGMGCGSSPCECHOEEDFRVTKXDI
      ORIPSLPSTQTL"
      109619..124347
      /note="Overlap with RP11-114N19, Accession AC007262"
      110216
      /note="217314: c; 114N19: t"
      /replace="t"
      110755
      /note="217314: g; 114N19: a"
      /replace="a"
      111115
      /note="217314: c; 114N19: g"
      /replace="g"
      114198
      /note="217314: g; 114N19: a"
      /replace="a"
      114690
      /note="217314: c; 114N19: t"
      /replace="t"
      115183
      /note="217314: t; 114N19: tatat"
      /replace="tatat"
      115210..115280
      /note="low quality data"
      116847
      /note="217314: t; 114N19: c"
      /replace="c"
      117131
      /note="217314: g; 114N19: a"
      /replace="a"
      117408
      /note="217314: a; 114N19: c"
      /replace="c"
      117633
      /note="217314: g; 114N19: t"
      /replace="t"

```

Chemistry: Big Dye terminators and primers
 Assembly program: Phrap; version 0.990399

variation	118901	/note="217314: a; 114N19: g"
	/replace="g"	
variation	118988	/note="217314: c; 114N19: t"
	/replace="t"	
variation	119250	/note="217314: a; 114N19: g"
	/replace="g"	
variation	119718	/note="217314: g; 114N19: t"
	/replace="t"	
variation	120914	/note="217314: c; 114N19: a"
	/replace="a"	
variation	121395..121413	/note="217314: caaaaaaaaaaaaaaa; 114N19: c"
	/replace="c"	
variation	123019	/note="217314: g; 114N19: a"
	/replace="a"	
variation	123721	/note="217314: c; 114N19: t"
	/replace="t"	
variation	123900	/note="217314: t; 114N19: c"
	/replace="c"	
variation	123926..123927	/note="217314: tt; 114N19: t"
	/replace="t"	
variation	124015	/note="217314: t; 114N19: c"
	/replace="c"	
variation	124182	/note="217314: g; 114N19: a"
	/replace="a"	
BASE COUNT	43520	a 24500 c 22595 g 33732 t
ORIGIN		

[illegible]

JOURNAL	Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 129837)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (25-FEB-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE	5 (bases 1 to 129837)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Mar 1, 1999 this sequence version replaced gi:3213168.

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapientewatson@wustl.edu
 Summary Statistics
 Center project name: H_PD0530J23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/G7/CH7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RPcl-4, prepared by Plietzer de Jong and coworkers at the Roswell Park Cancer Institute (<http://baopac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is GS1-67A24, 200 bp overlap.
Actual start of this clone is at base position 1 of RP4-530U23,
actual end is at 7592 of GS1-67A24.

FEATURES
source

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
chromosome="7"  
map=7q21.1-q21.3"  
clone_RP4-530J33"  
clone_lib="RPCI-4"  
  
1:1172  
repeat_region  
1:p16-family="MALR"  
1173..1541  
repeat_region  
fpc_family="MALR"  
1542..1962  
repeat_region
```

```

repeat_region      /rpl_family="L1"
1980. .2394
/rpl_family="L1"
repeat_region      2396. .2582
/rpl_family="Alu"
repeat_region      2595. .3425
/rpl_family="L1"
repeat_region      3426. .3736
/rpl_family="Alu"
repeat_region      3757. .3931
/rpl_family="L1"
repeat_region      3932. .4095
/rpl_family="L1"
repeat_region      4365. .4762
/rpl_family="Other"
repeat_region      5079. .5262
/rpl_family="Alu"
repeat_region      5332. .5358
/rpl_family="MER4-group"
5341. .5804
/rpl_family="MER4-group"
6123. .6217
/rpl_family="AT_rich"
6810. .6911
/rpl_family="AT_rich"
6932. .6960
/rpl_family="AT_rich"
6966. .7050
/rpl_family="AT_rich"
repeat_region      /rpl_family="MER73-group"
7086. .7423
/rpl_family="L2"
8602. .8649
/rpl_family="L2"
8608. .8687
/rpl_family="MIR"
8688. .8924
/rpl_family="Alu"
9006. .9146
/rpl_family="L2"
9147. .9927
/rpl_family="MER73-group"
9928. .10346
/rpl_family="L2"
11104. .11413
/rpl_family="Alu"
complement(12342. .12525)
/db_xref="GI:1113437"
13093. .13412
/rpl_family="L1"
14362. .14549
/rpl_family="L1"
14553. .14967
/rpl_family="L1"
15716. .15747
/rpl_family="AT_rich"
15786. .15821
/rpl_family="POLY_A"
17133. .17361
/rpl_family="MIR"
17455. .17841
/rpl_family="MALR"
17962. .18047
/rpl_family="L2"
18232. .18295
/rpl_family="L1"
18537. .18650
/rpl_family="MIR"
18651. .18943
/rpl_family="Alu"
18944. .18957
/rpl_family="MIR"
18958. .19010
/rpl_family="(TA)n"

```

```

repeat_region      19011. .19068
/rpl_family="MIR"
repeat_region      20552. .20660
/rpl_family="L2"
repeat_region      20891. .21251
/rpl_family="MALR"
repeat_region      22154. .22270
/rpl_family="L2"
repeat_region      22376. .22531
/rpl_family="L2"
repeat_region      22573. .22637
/rpl_family="AT_rich"
repeat_region      23327. .23702
/rpl_family="MER1_type?"
24527. .24650
/rpl_family="MER94"
24950. .25256
/rpl_family="Alu"
25634. .25677
/rpl_family="(CA)n"
25830. .25999
/rpl_family="MIR"
26413. .26449
/rpl_family="AT_rich"
26664. .26697
/rpl_family="(TA)n"
27087. .27107
/rpl_family="AT_rich"
29530. .29577
/rpl_family="G-rich"
33727. .33882
/rpl_family="MIR"

```

```

Query Match      0.6%; Score 20; DB 9; Length 129837;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 728 attcatggcatgctgcctga 747
|||||
Db 128831 ATTGATGGCATGCTGCTGA 128850

```

Search completed: March 25, 2002, 19:03:19
 Job time: 26263 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 13:24:21 ; Search time 97.89 Seconds
(without alignments)
7248.499 Million cell updates/sec

File: US-09-697-089-1
Perfect score: 3133
Sequence: 1 cgcctacgcccgcgtgcgaag.....aagtgactcgaacagcta 3133

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 351203 seqs, 113238999 residues

Word size : 0
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/2/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/2/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/2/lna/6C.COMB.seq:*
6: /cgn2_6/prodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	0.6	1094	2 US-08-870-518-34	Sequence 34, Appl
2	18	0.6	1515	3 US-08-747-221B-16	Sequence 16, Appl
3	18	0.6	1515	3 US-08-747-221B-17	Sequence 17, Appl
4	18	0.6	1515	4 US-09-005-051-16	Sequence 16, Appl
5	18	0.6	1515	4 US-09-005-051-17	Sequence 17, Appl
6	18	0.6	1611	6 5213972-6	Patent No. 5213972
7	18	0.6	1650	3 US-08-747-221B-60	Sequence 60, Appl
8	18	0.6	1650	3 US-08-747-221B-61	Sequence 61, Appl
9	18	0.6	1650	4 US-09-005-051-60	Sequence 60, Appl
10	18	0.6	1650	4 US-09-005-051-61	Sequence 61, Appl
11	18	0.6	1721	3 US-08-857-213-2	Sequence 2, Appl
12	18	0.6	1800	1 US-08-139-937-11	Sequence 11, Appl
13	18	0.6	1800	1 PCT-US93-11310-11	Sequence 11, Appl
14	18	0.6	1982	3 US-08-747-221B-13	Sequence 13, Appl
15	18	0.6	1982	3 US-08-747-221B-15	Sequence 15, Appl
16	18	0.6	1982	4 US-09-005-051-13	Sequence 13, Appl
17	18	0.6	1982	4 US-09-005-051-15	Sequence 15, Appl
18	18	0.6	2144	3 US-08-747-221B-57	Sequence 57, Appl
19	18	0.6	2144	3 US-08-747-221B-59	Sequence 59, Appl
20	18	0.6	2144	4 US-09-005-051-57	Sequence 57, Appl
21	18	0.6	2144	4 US-09-005-051-59	Sequence 59, Appl
22	18	0.6	3000	4 US-09-192-104-1	Sequence 1, Appl
23	18	0.6	3000	4 US-09-543-446-1	Sequence 1, Appl
24	18	0.6	4982	4 US-08-699-103B-1	Sequence 1, Appl
25	17	0.5	290	4 US-09-037-990B-27	Sequence 27, Appl
26	17	0.5	383	4 US-08-991-789A-171	Sequence 171, App
27	17	0.5	661	4 US-09-037-990B-8	Sequence 8, Appl

28	17	0.5	685	4 US-08-998-416-818	Sequence 818, App
29	17	0.5	1536	4 US-09-318-443-5	Sequence 5, Appl
30	17	0.5	1563	2 US-08-743-637B-183	Sequence 183, App
31	17	0.5	1682	4 US-09-318-443-7	Sequence 7, Appl
32	17	0.5	1665	4 US-09-178-252-26	Sequence 26, Appl
33	17	0.5	2035	4 US-08-960-780-10	Sequence 10, Appl
34	17	0.5	2035	4 US-09-073-898-10	Sequence 10, Appl
35	17	0.5	2521	4 US-09-115-446-1	Sequence 1, Appl
36	17	0.5	2526	4 US-09-115-446-5	Sequence 5, Appl
37	17	0.5	4165	1 US-08-095-737-1	Sequence 1, Appl
38	17	0.5	4165	1 US-08-480-145-1	Sequence 1, Appl
39	17	0.5	4165	2 US-08-477-389-1	Sequence 1, Appl
40	17	0.5	4437	1 US-08-559-303B-72	Sequence 72, Appl
41	17	0.5	4437	4 US-09-175-828-72	Sequence 72, Appl
42	17	0.5	441529	4 US-09-103-840A-1	Sequence 1, Appl
43	16	0.5	36	1 US-08-137-117D-48	Sequence 48, Appl
44	16	0.5	36	1 US-08-436-717-48	Sequence 48, Appl
45	16	0.5	84	4 US-08-943-731-153	Sequence 153, App
46	16	0.5	132	1 US-08-479-577-1	Sequence 1, Appl
47	16	0.5	132	2 US-08-756-977-1	Sequence 1, Appl
48	16	0.5	132	5 PCT-US93-05331-5	Sequence 5, Appl
49	16	0.5	284	3 US-08-621-018B-20	Sequence 20, Appl
50	16	0.5	300	3 US-08-171-385-26	Sequence 26, Appl
51	16	0.5	311	4 US-08-361-441B-40	Sequence 40, Appl
52	16	0.5	311	4 US-08-916-576B-40	Sequence 40, Appl
53	16	0.5	355	4 US-08-991-789A-78	Sequence 78, Appl
54	16	0.5	355	4 US-08-991-789A-28	Sequence 28, App
55	16	0.5	424	1 US-08-470-179-100	Sequence 100, App
56	16	0.5	434	4 US-08-943-731-127	Sequence 127, App
57	16	0.5	520	3 US-08-899-786-14	Sequence 14, Appl
58	16	0.5	562	3 US-08-899-786-16	Sequence 16, Appl
59	16	0.5	573	4 US-08-906-156A-67	Sequence 67, Appl
60	16	0.5	574	3 US-08-899-786-15	Sequence 15, Appl
61	16	0.5	590	4 US-08-906-156A-19	Sequence 19, Appl
62	16	0.5	603	4 US-09-385-982-233	Sequence 233, App
63	16	0.5	607	4 US-09-328-111-205	Sequence 205, App
64	16	0.5	625	4 US-09-385-982-101	Sequence 101, App
65	16	0.5	697	1 US-08-468-347-25	Sequence 25, Appl
66	16	0.5	697	1 US-08-467-389-25	Sequence 25, Appl
67	16	0.5	697	2 US-08-779-379-25	Sequence 25, Appl
68	16	0.5	697	2 US-08-469-219-25	Sequence 25, Appl
69	16	0.5	697	4 US-09-228-152-25	Sequence 25, Appl
70	16	0.5	732	4 US-08-916-576B-5	Sequence 5, Appl
71	16	0.5	733	1 US-08-152-485-3	Sequence 3, Appl
72	16	0.5	733	1 US-08-463-089-3	Sequence 3, Appl
73	16	0.5	733	1 US-08-461-360A-3	Sequence 3, Appl
74	16	0.5	733	1 US-08-461-359-3	Sequence 3, Appl
75	16	0.5	733	5 PCT-US94-12904-3	Sequence 3, Appl
76	16	0.5	795	2 US-08-822-028-11	Sequence 11, Appl
77	16	0.5	795	4 US-08-479-285-11	Sequence 11, Appl
78	16	0.5	803	3 US-08-454-928-9	Sequence 9, Appl
79	16	0.5	843	4 US-09-296-284-24	Sequence 24, Appl
80	16	0.5	897	5 PCT-US96-05320A-303	Sequence 303, App
81	16	0.5	921	4 US-09-296-284-3	Sequence 3, Appl
82	16	0.5	1047	2 US-08-535-276-7	Sequence 7, Appl
83	16	0.5	1054	1 US-08-152-485-1	Sequence 1, Appl
84	16	0.5	1054	1 US-08-463-089-1	Sequence 1, Appl
85	16	0.5	1054	1 US-08-461-360A-1	Sequence 1, Appl
86	16	0.5	1054	1 US-08-461-359-1	Sequence 1, Appl
87	16	0.5	1054	5 PCT-US94-12904-1	Sequence 1, Appl
88	16	0.5	1081	1 US-08-832-883-57	Sequence 57, Appl
89	16	0.5	1081	2 US-08-832-877-57	Sequence 57, Appl
90	16	0.5	1103	3 US-08-899-786-13	Sequence 13, Appl
91	16	0.5	1104	1 US-08-423-399B-36	Sequence 36, Appl
92	16	0.5	1134	4 US-08-432-994A-9	Sequence 9, Appl
93	16	0.5	1166	5 US-08-423-399B-34	Sequence 34, Appl
94	16	0.5	1166	5 PCT-US96-12198B-1	Sequence 1, Appl
95	16	0.5	1173	3 US-08-740-644-1	Sequence 1, Appl
96	16	0.5	1178	2 US-08-933-750C-88	Sequence 88, Appl
97	16	0.5	1178	2 US-09-234-613-88	Sequence 88, Appl
98	16	0.5	1244	2 US-08-204-288-3	Sequence 3, Appl
99	16	0.5	1284	2 US-07-882-790-3	Sequence 3, Appl
100	16	0.5	1330	2 US-08-868-288A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-870-518-34
Sequence 34, Application US/08870518
Patent No. 5925566
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
APPLICANT: Galcheva-Gargova, Zoya
TITLE OF INVENTION: NON-ACTIVATED RECEPTOR COMPLEX
TITLE OF INVENTION: PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,518
FILING DATE: 06-JUN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,219
FILING DATE: 06-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/102001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1094 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-870-518-34

Query Match 0.6%; Score 18; DB 2; Length 1094;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 766 ttctcctcttgatggct 783
|||||
Db 755 ttttcctcttgatggct 772

RESULT 2
US-08-747-221B-16/C
Sequence 16, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive

CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: PC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1515 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1515
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 298
US-08-747-221B-16

Query Match 0.6%; Score 18; DB 3; Length 1515;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
|||||
Db 363 ctccaattcagaatcc 346

RESULT 3
US-08-747-221B-17
Sequence 17, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington
 REGISTRATION NUMBER: 37,459
 REFERENCE/DOCKET NUMBER: FC-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 970/493-7272
 TELEFAX: 970/484-9505
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1515 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-747-221B-17

Query Match 0.6%; Score 18; DB 3; Length 1515;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1002 ctccaattcagaatcc 1019
 ||||||||||||||||
 1153 CTCCAATTCAGAAATCC 1170

RESULT 4
 US-09-005-051-16/c
 Sequence 16, Application US/09005051
 Patent No. 6291222

GENERAL INFORMATION:
 APPLICANT: Silver, Gary W.
 TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid
 NUMBER OF SEQUENCES: 66
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Carol Talkington Verser, Ph.D.
 STREET: 1825 Sharp Point Drive
 CITY: Fort Collins
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80525
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: Windows 95
 SOFTWARE: Wordperfect for Windows, Version 7.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/005,051
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/747,221
 FILING DATE: NO. 6291222e1, 12, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Verser, Carol Talkington
 REGISTRATION NUMBER: 37,459
 REFERENCE/DOCKET NUMBER: FC-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 970/493-7272
 TELEFAX: 970/484-9505
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1515 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1515
 FEATURE:

NAME/KEY: Asx = Asn or Asp
 LOCATION: 298
 US-09-005-051-16

Query Match 0.6%; Score 18; DB 4; Length 1515;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1002 ctccaattcagaatcc 1019
 ||||||||||||||||
 363 CTCCAATTCAGAAATCC 346

RESULT 5
 US-09-005-051-17
 Sequence 17, Application US/09005051
 Patent No. 6291222

GENERAL INFORMATION:
 APPLICANT: Silver, Gary W.
 TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid
 NUMBER OF SEQUENCES: 66
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Carol Talkington Verser, Ph.D.
 STREET: 1825 Sharp Point Drive
 CITY: Fort Collins
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80525
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: Windows 95
 SOFTWARE: Wordperfect for Windows, Version 7.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/005,051
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/747,221
 FILING DATE: NO. 6291222e1, 12, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Verser, Carol Talkington
 REGISTRATION NUMBER: 37,459
 REFERENCE/DOCKET NUMBER: FC-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 970/493-7272
 TELEFAX: 970/484-9505
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1515 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-09-005-051-17

Query Match 0.6%; Score 18; DB 4; Length 1515;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1002 ctccaattcagaatcc 1019
 ||||||||||||||||
 1153 CTCCAATTCAGAAATCC 1170

RESULT 6
 5213972-6/c
 Patent No. 5213972

APPLICANT: MCCANDLISS, RUSSELL J.; ANDERSON, DAVID M.
TITLE OF INVENTION: FERMENTATION PROCESS FOR THE PRODUCTION
OF PYRIMIDINE DEOXYRIBONUCLEOSIDES
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/448,158
FILING DATE: 08-DEC-1989
SEQ ID NO: 6:
LENGTH: 1611
5213972-6

Query Match 0.6%; Score 18; DB 6; Length 1611;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 343 aggattaaaggacttgc 360
|||||
979 AGGATTAAAGGACTTGT 962

RESULT 7

US-08-747-221B-60/c
Sequence 60, Application US/08747221B
Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1650 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1650

FEATURE:

NAME/KEY: ASX - Asn or Asp

LOCATION: 433

US-08-747-221B-60

Query Match 0.6%; Score 18; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1002 ctccaattcagaatcc 1019
|||||
DB 438 CTCCAATTCAAGAAATCC 481

RESULT 8

US-08-747-221B-61
Sequence 61, Application US/08747221B
Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:

LENGTH: 1650 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-747-221B-61

Query Match 0.6%; Score 18; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctccaattcagaatcc 1019
|||||
DB 1153 CTCCAATTCAAGAAATCC 1170

RESULT 9

US-09-005-051-60/c
Sequence 60, Application US/09005051
Patent No. 6291222

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1650
FEATURE:
NAME/KEY: Asx - Asn or Asp
LOCATION: 433
US-09-005-051-60

Query Match 0.6%; Score 18; DB 4; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 ctccaattcagaatcc 1019
Db 498 CTCCAATTGAGAAATCC 481

RESULT 10
US-09-005-051-61
Sequence 61, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-005-051-61

Query Match 0.6%; Score 18; DB 4; Length 1650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 ctccaattcagaatcc 1019
Db 1153 CTCCAATTGAGAAATCC 1170

RESULT 11
US-08-857-213-2/C
Sequence 2, Application US/08857213
Patent No. 6054290
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/857,213
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0297 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBRNGT01
CLONE: 148415
US-08-857-213-2

Query Match 0.6%; Score 18; DB 3; Length 1721;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 atgaattcataaggac 53
|||||
Db 1538 ATGAATTCATTAAGAC 1521

USUT 12
-08-139-937-11
Sequence 11, Application US/08139937
Patent No. 5821070

GENERAL INFORMATION:

APPLICANT: LEE, WEN-HWA
TITLE OF INVENTION: CELLULAR GENES ENCODING
TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/139,937
FILING DATE: 20-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/979,156
FILING DATE: 20-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9370
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-139-937-11

Query Match 0.6%; Score 18; DB 1; Length 1800;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2831 aaagaacctgaaaaa 2848
|||||
Db 307 AAAGAACCCTCGAAAAA 324

RESULT 13.

PCT-US93-11310-11
Sequence 11, Application PC/TUS9311310

GENERAL INFORMATION:

APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
TITLE OF INVENTION: CELLULAR GENES ENCODING
TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11310
FILING DATE: 19-NOV-1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9790
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US93-11310-11

Query Match 0.6%; Score 18; DB 5; Length 1800;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2831 aaagaacctgaaaaa 2848
|||||
Db 307 AAAGAACCCTCGAAAAA 324

RESULT 14

US-08-747-221B-13/C

Sequence 13, Application US/08747221B

Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.
TITLE OF INVENTION: No. 6063610e1 Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1517
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 300
US-08-747-221B-13

Query Match 0.6%; Score 18; DB 3; Length 1982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 ctccaatcagaatcc 1019
|||||

Db 365 CTCCAATTCAGAAATCC 348

RESULT 15
US-08-747-221B-15
Sequence 15, Application US/08747221B
Patent No. 6063610
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,221B
FILING DATE: No. 6063610ember 12, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-747-221B-15

Query Match 0.6%; Score 18; DB 3; Length 1982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 ctccaatcagaatcc 1019
|||||

Db 1618 CTCCAATTCAGAAATCC 1635

RESULT 16
US-09-005-051-13/C
Sequence 13, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1982 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1517
FEATURE:
NAME/KEY: Asx = Asn or Asp
LOCATION: 300
US-09-005-051-13

Query Match 0.6%; Score 18; DB 4; Length 1982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 ctccaatcagaatcc 1019

Db 365 CTCCAATTGAGAAATCC 348

RESULT 17

US-09-005-051-15

Sequence 15, Application US/09005051

Patent No. 6291222

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/005,051

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/747,221

FILING DATE: No. 6291222el December 12, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 1982 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

-09-005-051-15

Query Match 0.6%; Score 18; DB 4; Length 1982;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019

Db 1618 CTCCAATTGAGAAATCC 1635

RESULT 18

US-08-747-221B-57/c

Sequence 57, Application US/08747221B

Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610el December 12, 1996

CLASSIFICATION: 435

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610el December 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 2144 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 30..1682

FEATURE:

NAME/KEY: Asx = Asn or Asp

LOCATION: 462

US-08-747-221B-57

Query Match 0.6%; Score 18; DB 3; Length 2144;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019

Db 527 CTCCAATTGAGAAATCC 510

RESULT 19

US-08-747-221B-59

Sequence 59, Application US/08747221B

Patent No. 6063610

GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carol Talkington Verser, Ph.D.

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: Wordperfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610el December 12, 1996

CLASSIFICATION: 435

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2144 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-747-221B-59

Query Match 0.6%; Score 18; DB 3; Length 2144;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1002 ctccaattcagaatcc 1019
1618 CTCCAATTGAGAAATCC 1635

RESULT 20
US-09-005-051-57/C
; Sequence 57, Application US/09005051
; Patent No. 6291222
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Misnewski, Nancy
; TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,051
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/747,221
; FILING DATE: No. 6291222ember 12, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2144 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1682
```

```

; FEATURE:
; NAME/KEY: Asx = Asn or Asp
; LOCATION: 462
; US-09-005-051-57

Query Match 0.6%; Score 18; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctccaattcagaatcc 1019
DB 527 CTCCAATTGAGAAATCC 510

RESULT 21
US-09-005-051-59
; Sequence 59, Application US/09005051
; Patent No. 6291222
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Misnewski, Nancy
; TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,051
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/747,221
; FILING DATE: No. 6291222ember 12, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2144 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-09-005-051-59

Query Match 0.6%; Score 18; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctccaattcagaatcc 1019
DB 1618 CTCCAATTGAGAAATCC 1635

RESULT 22
US-09-192-104-1
```

```
Sequence 1, Application US/09192104B
Patent No. 6184020
GENERAL INFORMATION:
APPLICANT: Alexander Blankovsky
APPLICANT: Tony Byun
APPLICANT: Alan V. Klotz
APPLICANT: Alan Sloma
APPLICANT: Maria Tang
APPLICANT: Mikio Fujii
APPLICANT: Chigusa Marumoto
APPLICANT: Lene Venke Kofod
TITLE OF INVENTION: Polypeptides Having Aminopeptidase
Activity And Nucleic Acids Encoding Same
FILE REFERENCE: 5379.200-US
CURRENT APPLICATION NUMBER: US/09/192,104B
EARLIER FILING DATE: 1998-11-13
EARLIER APPLICATION NUMBER: 60/069719
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: 1465/97
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: PA 1998 00670
EARLIER FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3000
TYPE: DNA
ORGANISM: Sphingomonas
US-09-192-104-1
```

```
Query Match 0.6%; Score 18; DB 4; Length 3000;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2127 gctgtgtgtgtgtgaagc 2144
|||||
Db 1911 gctgtgtgtgtgtgaagc 1928
```

```
RESULT 23
US-09-543-446-1
Sequence 1, Application US/09543446
Patent No. 6303360
GENERAL INFORMATION:
APPLICANT: Alexander Blankovsky
APPLICANT: Tony Byun
APPLICANT: Alan V. Klotz
APPLICANT: Alan Sloma
APPLICANT: Maria Tang
APPLICANT: Mikio Fujii
APPLICANT: Chigusa Marumoto
APPLICANT: Lene Venke Kofod
TITLE OF INVENTION: Polypeptides Having Aminopeptidase
Activity And Nucleic Acids Encoding Same
FILE REFERENCE: 5379.210-US
CURRENT APPLICATION NUMBER: US/09/543,446
EARLIER FILING DATE: 2000-04-05
EARLIER APPLICATION NUMBER: 60/069719
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: 1465/97
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: PA 1998 00670
EARLIER FILING DATE: 1998-05-15
EARLIER APPLICATION NUMBER: 09/192,104
EARLIER FILING DATE: 1998-11-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3000
TYPE: DNA
ORGANISM: Sphingomonas
US-09-543-446-1
```

```
Query Match 0.6%; Score 18; DB 4; Length 3000;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2127 gctgtgtgtgtgtgaagc 2144
|||||
Db 1911 gctgtgtgtgtgtgaagc 1928
```

```
RESULT 24
US-08-699-103B-1/c
Sequence 1, Application US/08699103B
Patent No. 6107462
GENERAL INFORMATION:
APPLICANT: Rine, Jasper D.
APPLICANT: Hampton, Randolph
TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
CHOLESTEROL SYNTHESIS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/699,103B
FILING DATE: 16-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,581
FILING DATE: 17-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 09272/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/322-5070
TELEFAX: 650/854-0875
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-699-103B-1
```

```
Query Match 0.6%; Score 18; DB 3; Length 4982;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 391 ccctgtgtgaagatattg 408
|||||
Db 1743 CCCTGTGTGAAGATATTG 1726
```

```
RESULT 25
US-09-037-990B-27
Sequence 27, Application US/09037990B
Patent No. 6248519
GENERAL INFORMATION:
APPLICANT: ENGEL, Stacia R.
DESCENZO, Richard A.
MORENZONI, Richard A.
US-09-037-990B-27
```

```

: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/991,789A
: FILING DATE: 11-Dec-1997
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Potler, Jane E. R.
: REGISTRATION NUMBER: 33,332
: REFERENCE/DOCKET NUMBER: 210121.419C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 171:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 383 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: SEQUENCE DESCRIPTION: SEQ ID NO: 171:
:
: US-08-991-789A-171
:
:
: Query Match 0.5%; Score 17; DB 4; Length 383;
: Best Local Similarity 100.0%; Pred. No. 1.4e+02;
: Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0
:
: QY 409 acattattttaactgt 425
: |||||
: Db 35 ACATTATTTTACTCTG 19
:
: RESULT 27
: US-09-037-990B-8
: Sequence 8, Application US/09037990B
: Patent No. 6248519
:
: GENERAL INFORMATION:
: APPLICANT: ENGEL, Stacia R.
: DESCENZO, Richard A.
: MORENZONI, Richard A.
: IRELAN, Nancy A.
:
: TITLE OF INVENTION: DETECTION OF FERMENTATION-RELATED
: MICROORGANISMS
:
: NUMBER OF SEQUENCES: 100
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/037,990B
: FILING DATE: 11-Mar-1999
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: <Unknown>
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Sharp, Jeffrey S.
: REGISTRATION NUMBER: 31,879
: REFERENCE/DOCKET NUMBER: 29520/30001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELEX: <Unknown>
:
: INFORMATION FOR SEQ ID NO: 8:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 661 base pairs

```



```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 183:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1563 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Klebsiella pneumoniae
; US-08-743-637B-183
```

```
Query Match 0.5%; Score 17; DB 2; Length 1563;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 505 tccgcagctctcag 521
|||||
74 TCCGCGAGGCTCTTCAG 58
```

```
RESULT 31
US-09-318-443-7/C
; Sequence 7, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briyanolou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1682
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-318-443-7
```

```
Query Match 0.5%; Score 17; DB 4; Length 1682;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
1374 caggagtagacacagcag 1390
|||||
1501 CAGGAGTAGACAGCAGCAG 1485
```

```
RESULT 32
US-09-178-252-26
; Sequence 26, Application US/09178252
; Patent No. 6218188
; GENERAL INFORMATION:
; APPLICANT: Cardineau, Guy A.
; APPLICANT: Steiman, Steven J.
; APPLICANT: Naraya, Kenneth E.
; TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Toxins
; FILE REFERENCE: MA-714XC2
; CURRENT APPLICATION NUMBER: US/09/178,252
; CURRENT FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 60/065,215
; EARLIER FILING DATE: 1997-11-12
; EARLIER APPLICATION NUMBER: 60/076,445
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
```

```
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic B.t. toxin gene
; US-09-178-252-26
```

```
Query Match 0.5%; Score 17; DB 4; Length 1965;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 463 accaacacacacaccgc 479
|||||
Db 1519 accaacacacacaccgc 1535
```

```
RESULT 33
US-08-960-780-10
; Sequence 10, Application US/08960780
; Patent No. 6204435
; GENERAL INFORMATION:
; APPLICANT: Feltelson, Gerald S.
; APPLICANT: Schneck, H. Ernest
; APPLICANT: Naraya, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeltz, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 158C2-pt1
; US-08-960-780-10
```

```
Query Match 0.5%; Score 17; DB 4; Length 2035;
```

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2620 aagatggaatgaagct 2636
|||||
Db 1973 AAGATGGAATGAAGCT 1989

RESULT 34
US-09-073-898-10

; Sequence 10, Application US/09073898
; Patent No. 6242669

; GENERAL INFORMATION:

; APPLICANT: Feltelson, Jerald S.

; APPLICANT: Schaefer, H. Ernest

; APPLICANT: Narva, Kenneth E.

; APPLICANT: Stockhoff, Brian A.

; APPLICANT: Schmeits, James

; APPLICANT: Loefer, David

; APPLICANT: Dullum, Charles Joseph

; APPLICANT: Muller-Cohn, Judy

; APPLICANT: Stamp, Lisa

; APPLICANT: Morill, George

; TITLE OF INVENTION: No. 624269e1 Pesticidal Toxins and Nucleotide

; NUMBER OF SEQUENCES: 144

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Saliwanichik, Lloyd & Saliwanichik

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: US

; ZIP: 32606-6669

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/073.898

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/029,848

; FILING DATE: 30-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/960,780

; FILING DATE: 30-OCT-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Sanders, Jay M.

; REGISTRATION NUMBER: 39,355

; REFERENCE/DOCKET NUMBER: MA-708C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 352-375-8100

; TELEFAX: 352-372-5800

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2035 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; INDIVIDUAL ISOLATE: 158C2-PT1

US-09-073-898-10

Query Match 0.5%; Score 17; DB 4; Length 2035;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2620 aagatggaatgaagct 2636

|||||
Db 1973 AAGATGGAATGAAGCT 1989

RESULT 35
US-09-115-446-1/C

; Sequence 1, Application US/09115446
; Patent No. 6165719

; GENERAL INFORMATION:

; APPLICANT: Chandry, George K.

; APPLICANT: Gargus, Jay J.

; APPLICANT: Gutman, George

; APPLICANT: Fantino, Emmanuelle

; APPLICANT: Kaiman, Katarin

; TITLE OF INVENTION: hKCA3/KCNN3 SMALL CONDUCTANCE CALCIUM

; TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC

; FILE REFERENCE: 07306/014001

; CURRENT APPLICATION NUMBER: US/09/115.446

; CURRENT FILING DATE: 1998-07-14

; EARLIER APPLICATION NUMBER: 60/052,556

; EARLIER FILING DATE: 1997-07-15

; EARLIER APPLICATION NUMBER: 60/070,741

; EARLIER FILING DATE: 1998-01-08

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 2521

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (287)...(2479)

US-09-115-446-1

Query Match 0.5%; Score 17; DB 4; Length 2521;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 tgcctggaagctgag 755
|||||
Db 492 tgcctggaagctgag 476

RESULT 36

US-09-115-446-5/C

; Sequence 5, Application US/09115446
; Patent No. 6165719

; GENERAL INFORMATION:

; APPLICANT: Chandry, George K.

; APPLICANT: Gargus, Jay J.

; APPLICANT: Gutman, George

; APPLICANT: Fantino, Emmanuelle

; APPLICANT: Kaiman, Katarin

; TITLE OF INVENTION: hKCA3/KCNN3 SMALL CONDUCTANCE CALCIUM

; TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC

; FILE REFERENCE: 07306/014001

; CURRENT APPLICATION NUMBER: US/09/115.446

; CURRENT FILING DATE: 1998-07-14

; EARLIER APPLICATION NUMBER: 60/052,556

; EARLIER FILING DATE: 1997-07-15

; EARLIER APPLICATION NUMBER: 60/070,741

; EARLIER FILING DATE: 1998-01-08

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 2526

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-115-446-5

Query Match 0.5%; Score 17; DB 4; Length 2526;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 tgcctcgaagctgcg 755
|||||
Db 497 TGCCTGAGAGCTGCCG 481

RESULT 37

US-08-095-737-1/C
; Sequence 1, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: Diflore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,737
; FILING DATE: 19930722
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..2709
US-08-095-737-1

Query Match 0.5%; Score 17; DB 1; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2811 atttagtcatttt 2827
|||||
Db 4070 ATTTAGTGCAATTTT 4054

RESULT 38

US-08-480-145-1/C
; Sequence 1, Application US/08480145
; Patent No. 5717067
; GENERAL INFORMATION:
; APPLICANT: Diflore, Pier P

APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,145
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,737
; FILING DATE: 22-JUL-1993

ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A
; REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH060.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..2709

US-08-480-145-1

Query Match 0.5%; Score 17; DB 1; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2811 atttagtcatttt 2827
|||||
Db 4070 ATTTAGTGCAATTTT 4054

RESULT 39

US-08-477-389-1/C
; Sequence 1, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:
; APPLICANT: Diflore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; TITLE OF INVENTION: Factor Receptor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477.389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/095.737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 21..2709
US-08-477-389-1

Query Match 0.5%; Score 17; DB 2; Length 4165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2811 atttagtgatctttt 2827
|||||

Db 4070 attttaggtgcatTTTT 4054

RESULT 40
US-08-559-303B-72
Sequence 72, Application US/08559303B
Patent No. 5624501
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559.303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-559-303B-72

Query Match 0.5%; Score 17; DB 1; Length 4437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3052 aatttgatgatgatgat 3068
|||||

Db 943 aatttgatgatgatgat 959

RESULT 41
US-09-175-828-72
Sequence 72, Application US/09175828
Patent No. 6221643
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BLOOM'S SYNDROME
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/175.828
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/559.303
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES

ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-175-828-72

Query Match 0.5%; Score 17; DB 4; Length 4437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3052 aatttgatgatgatg 3068
|||||
DB 943 AATTGATGATGATGAT 959

RESULT 42
US-09-103-840A-1/C
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 441529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 0.5%; Score 17; DB 4; Length 441529;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2400 aacctgaagaagatg 2416
|||||
DB 3725573 AACCTGAGAGATGTG 3725557

RESULT 43
US-08-137-117D-48/C
Sequence 48, Application US/08137117D
Patent No. 5795965
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, Masayuki
APPLICANT: SATO, Koh
APPLICANT: BENDIG, Mary
APPLICANT: JONES, Steven
APPLICANT: SALDANHA, Jose
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
NUMBER OF SEQUENCES: 158
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/137,117D
FILING DATE: 20-DEC-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NO PCT/JP92/00544
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-32084
FILING DATE: 19-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-95476
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/126/AOK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-137-117D-48

Query Match 0.5%; Score 16; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2182 ctctcgtggtgaagc 2197
|||||
DB 22 CTCCTCATGTGTGAAGC 7

RESULT 44
US-08-436-717-48/C
Sequence 48, Application US/08436717
Patent No. 5817790
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, Masayuki
APPLICANT: SATO, Koh
APPLICANT: BENDIG, Mary
APPLICANT: JONES, Steven
APPLICANT: SALDANHA, Jose
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
NUMBER OF SEQUENCES: 158
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,717
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/137,117

FILING DATE: 20-DEC-1993
APPLICATION NUMBER: WO PCT/JP92/00544
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-32084
FILING DATE: 19-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-95476
FILING DATE: 25-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/126/AAOK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-717-48

Query Match 0.5%; Score 16; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2182 ctcctcgtgtggaagc 2197
|||||

DB 22 CTCCTCATGTGTGAAC 7

RESULT 45
US-08-943-731-153/c
Sequence 153, Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCKOP, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTANTINOS D.
APPLICANT: SEREDA, LARISSA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACEK, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORRKO, JARMO
APPLICANT: ALA-KORRKO, LEENA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-153

Query Match 0.5%; Score 16; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1419 catgagccagagag 1434
|||||

DB 61 CATGAGCCAGAGAG 46

RESULT 46
US-08-479-577-1/c
Sequence 1, Application US/08479577
Patent No. 5633449
GENERAL INFORMATION:
APPLICANT: MILTON ZAITLIN, and Peter Palukaitis
TITLE OF INVENTION: Induction Of Resistance To
TITLE OF INVENTION: Viral Diseases In Plants
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Yahwak & Associates
STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,577
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/197,096
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-1123NM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
US-08-479-577-1

Query Match 0.5%; Score 16; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2541 gcagtgaaatcctag 2556
|||||
DB 100 GCAGTGAAATCCTAG 85

RESULT 47
US-08-756-977-1/c
; Sequence 1, Application US/08756977
; Patent No. 5945581

GENERAL INFORMATION:
APPLICANT: Zaitlin, Milton
APPLICANT: Palukaitis, Peter
TITLE OF INVENTION: INDUCTION OF RESISTANCE TO VIRAL
TITLE OF INVENTION: DISEASES IN PLANTS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/756,977
FILING DATE: 26-NOV-1996
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/10186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600

INFORMATION FOR SEQ. ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA (genomic)
US-08-756-977-1

Query Match 0.5%; Score 16; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2541 gcagtgaaatcctag 2556
|||||
DB 100 GCAGTGAAATCCTAG 85

RESULT 48
PCT-US93-05331-5/c
; Sequence 5, Application PC/TUS9305331
; GENERAL INFORMATION:

APPLICANT: Cornell Research Foundation, Inc.
TITLE OF INVENTION: Induction Of Resistance To
TITLE OF INVENTION: Virus Diseases By Transformation Of Plants With A Replicase Pc
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Yahwak & Associates.

STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05331
FILING DATE: 19930603
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: George W. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-1123
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951

INFORMATION FOR SEQ. ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US93-05331-5

Query Match 0.5%; Score 16; DB 5; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2541 gcagtgaaatcctag 2556
|||||
DB 100 GCAGTGAAATCCTAG 85

RESULT 49
US-08-621-018B-20
; Sequence 20, Application US/08621018B
; Patent No. 6060239

GENERAL INFORMATION:

APPLICANT: Stuart, Susan G.
APPLICANT: Hawkins, Phillip R.
APPLICANT: Sellhammer, Jeffrey J.
APPLICANT: Murry, Lynn E.

TITLE OF INVENTION: CELLUBREVIN HOMOLOGS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/621,018B

FILING DATE: March 22, 1996
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/409,373
FILING DATE: March 23, 1995

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0029-1 CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: SYNORAT01
CLONE: 370165
US-08-621-018B-20

Query Match 0.5%; Score 16; DB 3; Length 284;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2186 catggtggaagcagct 2201
|||||
Db 43 CATGCTGAGACCACT 58

RESULT 50
US-08-171-385-26
Sequence 26, Application US/08171385
Patent No. 5527884
GENERAL INFORMATION:
APPLICANT: Mary E. Russell
APPLICANT: Ulrike Utans
TITLE OF INVENTION: Mediators of Chronic Allograft
TITLE OF INVENTION: Rejection
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/171.385
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/006001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 300
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-171-385-26

Query Match 0.5%; Score 16; DB 1; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1995 aactggaagcagaat 2010
|||||
Db 282 AACTGAGACGAGCAAT 297

RESULT 51
US-08-361-441B-26
Sequence 26, Application US/08361441B
Patent No. 6077948
GENERAL INFORMATION:
APPLICANT: Russell, Mary E.
APPLICANT: Utans, Ulrike
TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/361.441B
FILING DATE: 21-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/171.385
FILING DATE: 21-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/014001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-361-441B-26

Query Match 0.5%; Score 16; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1995 aactggaagcagaat 2010
|||||
Db 282 AACTGAGACGAGCAAT 297

RESULT 52
US-08-916-576B-40
Sequence 40, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600
 CITY: WASHINGTON
 STATE: DC
 COUNTRY: US
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/916,576B
 FILING DATE:
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/024,347
 FILING DATE: 23-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: STEFFE, ERIC K.
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.0500001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 311 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 US-08-916-576B-40

Query Match 0.5%; Score 16; DB 4; Length 311;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1820 attgactctttaa 1835
 Db 108 ATTGACTCTTGA 123

RESULT 53
 US-08-991-789A-78
 Sequence 78, Application US/08991789A
 Patent No. 6225054
 GENERAL INFORMATION:
 APPLICANT: Fridakis, Tony N.
 Smith, John M.
 Reed, Steven G.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 TREATMENT AND DIAGNOSIS OF BREAST CANCER
 NUMBER OF SEQUENCES: 292
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed IP Law Group
 STREET: 701 Fifth Avenue, Suite 6300
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,789A
 FILING DATE: 11-Dec-1997
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Potler, Jane E. R.
 REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 78:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 355 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 78:
 US-08-991-789A-78

Query Match 0.5%; Score 16; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2557 ctcaaatcttcaaa 2572
 Db 149 CTCGAATCTTCACAA 164

RESULT 54
 US-08-991-789A-248
 Sequence 248, Application US/08991789A
 Patent No. 6225054
 GENERAL INFORMATION:
 APPLICANT: Fridakis, Tony N.
 Smith, John M.
 Reed, Steven G.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 TREATMENT AND DIAGNOSIS OF BREAST CANCER
 NUMBER OF SEQUENCES: 292
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed IP Law Group
 STREET: 701 Fifth Avenue, Suite 6300
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,789A
 FILING DATE: 11-Dec-1997
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Potler, Jane E. R.
 REGISTRATION NUMBER: 33,332
 REFERENCE/DOCKET NUMBER: 210121.419C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 248:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 355 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 248:
 US-08-991-789A-248

Query Match 0.5%; Score 16; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2557 ctcaaatcttcaaa 2572
 ||||||||||||||||

Db 149 CTCGATCTTCACAA 164

RESULT 55
US-08-470-179-100
; Sequence 100, Application US/08470179
; Patent No. 5645984
; GENERAL INFORMATION:
; APPLICANT: Huang Ph.D, Mai Mun
; TITLE OF INVENTION: Method and Compositions for
; TITLE OF INVENTION: Identification of Species in a Sample
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Trask, Britt and Rossa
; STREET: P.O. Box 2550
; CITY: Salt Lake City
; STATE: Utah
; COUNTRY: USA
; ZIP: 84110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/470.179
APPLICATION NUMBER: US/08/470.179
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sweigert Ph.D, Susan E.
REGISTRATION NUMBER: 36,289
REFERENCE/DOCKET NUMBER: 2601
TELECOMMUNICATION INFORMATION:
TELEPHONE: 801-532-1922
TELEFAX: 801-531-9168
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "gyra gene segment"
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Treponema pallidum
US-08-470-179-100

Query Match 0.5%; Score 16; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 tgggaagcttcaccc 29
|||||

Db 107 TGGGAGCTTCATCC 122

RESULT 56
US-08-943-731-127
; Sequence 127, Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:
; APPLICANT: PROCKOP, DARWIN J.
; APPLICANT: SPOTILA, LORETTA D.
; APPLICANT: DETRAS, CONSTANTINOS D.
; APPLICANT: SEREDA, LARISSA
; APPLICANT: LARSON, ANDREA W.
; APPLICANT: PACK, MICHAEL
; APPLICANT: COLIGE, ALAIN
; APPLICANT: EARLY, JAMES
; APPLICANT: KORRKO, JARMO

APPLICANT: ALA-KORRKO, LEENA, et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STATE: FLR.
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/943.731
APPLICATION NUMBER: US/08/943.731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-127

Query Match 0.5%; Score 16; DB 4; Length 434;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1979 tgatccttgctctc 1994
|||||

Db 196 TGTATCTTGTCTCTC 211

RESULT 57
US-08-899-786-14/c
; Sequence 14, Application US/08899786
; Patent No. 6001572
; GENERAL INFORMATION:
; APPLICANT: Toothman, Penelope
; TITLE OF INVENTION: Method of Identifying Aloe Using
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS


```

; SOFTWARE: wordperfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,786
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/022,611
; FILING DATE: 26-JULY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: UNI.07
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 520 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-899-786-14

Query Match      0.5%; Score 16; DB 3; Length 520;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 gattggtcaggatt 348
   |||
DB 98 GATTGGCTCAGAGTT 83

RESULT 58
US-08-899-786-16/c
; Sequence 16, Application US/08899786
; Patent No. 6001572
; GENERAL INFORMATION:
; APPLICANT: Toothman, Penelope
; TITLE OF INVENTION: Method of Identifying Aloe Using
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: wordperfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,786
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/022,611
; FILING DATE: 26-JULY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: UNI.07
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-899-786-16
```

```

; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-899-786-16

Query Match      0.5%; Score 16; DB 3; Length 562;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 gattggtcaggatt 348
   |||
DB 100 GATTGGCTCAGAGTT 85

RESULT 59
US-08-906-156A-67/c
; Sequence 67, Application US/08906156A
; Patent No. 6287854
; GENERAL INFORMATION:
; APPLICANT: SPURR, NIGEL K
; APPLICANT: GRAY, IAN C
; APPLICANT: STEWART, LORNA M
; TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHAYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,156A
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,655
; FILING DATE: 02-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,147
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,840
; FILING DATE: 23-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/96GB/02588
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1090-14
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: human
; US-08-906-156A-67
```

Query Match 0.5%; Score 16; DB 4; Length 573;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 tcattcacatgatattt 211
|||||
Db 76 TCATTACATGATTTT 61

RESULT 60
US-08-899-786-15/c

; Sequence 15, Application us/08899786
; Patent No. 6001572

; GENERAL INFORMATION:

; APPLICANT: Toothman, Penelope

; TITLE OF INVENTION: Method of Identifying Aloe Using

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Avenue, Suite 200

CITY: Englewood

STATE: Colorado

COUNTRY: USA

ZIP: 80111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB

COMPUTER: IBM pc compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,786

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/022,611

FILING DATE: 26-JULY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson

REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: UNI. 07

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433

INFORMATION FOR SEQ. ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 574 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE:

US-08-899-786-15

Query Match 0.5%; Score 16; DB 3; Length 574;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 333 gatttgctcagatt 348
|||||
Db 112 GATTGGCTCAGATT 97

RESULT 61
US-08-906-156A-19/c

; Sequence 19, Application US/08906156A
; Patent No. 6287854

; GENERAL INFORMATION:

; APPLICANT: SPURR, NIGEL K

; APPLICANT: GRAY, IAN C

; APPLICANT: STEWART, LORNA M

; TITLE OF INVENTION: DIAGNOSIS OF SUSCEPTIBILITY TO CANCER

; TITLE OF INVENTION: AND TREATMENT THEREOF

; NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/906,156A

FILING DATE: 05-AUG-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/042,655

FILING DATE: 02-APR-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/005,840

FILING DATE: 23-OCT-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/96GB/02588

FILING DATE: 22-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 1090-14

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-816-4000

TELEFAX: 703-816-4100

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 590 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Part of gene corresponding to IMAGE 264611

US-08-906-156A-19

Query Match 0.5%; Score 16; DB 4; Length 590;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 tcattcacatgatattt 211
|||||
Db 76 TCATTACATGATTTT 61

RESULT 62
US-09-385-982-233

; Sequence 233, Application US/09385982
; Patent No. 6262334

; GENERAL INFORMATION:

; APPLICANT: ENDEGE, WILSON O., ET AL.

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; TITLE OF INVENTION: PRODUCTS: II

; FILE REFERENCE: CCDNA-260XX

; CURRENT APPLICATION NUMBER: US/09/385,982

; EARLIER FILING DATE: 1999-08-30

; EARLIER APPLICATION NUMBER: 09/328,111

; EARLIER FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: 60/117,393

; EARLIER FILING DATE: 1999-01-27

; EARLIER APPLICATION NUMBER: 60/098,639

Tue Mar 26 11:35:52 2002

```
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 233
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(603)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-233
```

```
Query Match
Best Local Similarity 100.0%; Score 16; DB 4; Length 603;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 2412 atgcttatttcattc 2427

Db 373 atgcttatttcattc 388

RESULT 63

```
US-09-328-111-205/c
; Sequence 205, Application US/09328111
```

```
; GENERAL INFORMATION:
; PATENT NO. 6262333
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 205
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(607)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-205
```

```
Query Match
Best Local Similarity 100.0%; Score 16; DB 4; Length 607;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 2977 cagcattagtcagaaa 2992

Db 325 CAGCATTAGTCAGAAA 310

RESULT 64

```
US-09-385-982-101/c
; Sequence 101, Application US/09385982
; Patent No. 6262334
```

```
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 625
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(625)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-101
```

```
Query Match
Best Local Similarity 100.0%; Score 16; DB 4; Length 625;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 763 ttcttctctcttga 778

Db 159 ttcttctctcttga 144

RESULT 65

```
US-08-468-347-25/c
; Sequence 25, Application US/08468347
```

```
; GENERAL INFORMATION:
; PATENT NO. 5783421
; APPLICANT: Zeelon, Elisha P.
; APPLICANT: Werber, Moshe M.
; APPLICANT: Levanon, Avigdor
; TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,347
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/225,442
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0317/43020-A/JPM/ENB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-664-0525
; TELEFAX: 212-664-0525
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-468-347-25

Query Match 0.5%; Score 16; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
|||||
Db 678 TTTTGTGACTTAGTA 663

RESULT 66
US-08-467-389-25/C
Sequence 25, Application US/08467389
Patent No. 5824641
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,389
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-467-389-25

Query Match 0.5%; Score 16; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
|||||
Db 678 TTTTGTGACTTAGTA 663

RESULT 67
US-08-779-379-25/C
Sequence 25, Application US/08779379
Patent No. 5858970
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/779,379
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-779-379-25

Query Match 0.5%; Score 16; DB 2; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
|||||
Db 678 TTTTGTGACTTAGTA 663

RESULT 68
US-08-469-219-25/C
Sequence 25, Application US/08469219
Patent No. 5863534
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa

TITLE OF INVENTION: INHIBITORY ACTIVITY
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,219
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 697 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-469-219-25

Query Match 0.5%; Score 16; DB 2; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
Db 678 tttttgacttagta 663

SULT 69
US-09-228-152-25/C
Sequence 25, Application US/09228152
Patent No. 6211341
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Werber, Moshe M.
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa INHIBITORY ACTIVITY
FILE REFERENCE: 43020aYa
CURRENT APPLICATION NUMBER: US/09/228,152
CURRENT FILING DATE: 1999-01-11
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 25
LENGTH: 697
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA sequence
US-09-228-152-25

Query Match 0.5%; Score 16; DB 4; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2940 tttttgacttagta 2955
Db 678 tttttgacttagta 663

RESULT 70
US-08-916-576B-5/C
Sequence 5, Application US/08916576B
Patent No. 6171816
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: DILLON, PATRICK J.
APPLICANT: EBNER, REINHARD
APPLICANT: ENDRESS, GREGORY A.
TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,576B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/024,347
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0500001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 49..546
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 118..546
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 49..117
US-08-916-576B-5

Query Match 0.5%; Score 16; DB 4; Length 732;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1820 attgactcttgaa 1835
Db 589 attgactcttgaa 574

```
RESULT 71
US-08-152-485-3
; Sequence 3, Application US/08152485
; Patent No. 5539094
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,485
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-152-485-3

Query Match      0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2732 gttgaacatttgag 2747
|||||
Db 229 GTTGAACATTGGAG 244

RESULT 72
US-08-463-089-3
; Sequence 3, Application US/08463089
; Patent No. 5641866
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,089
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-463-089-3

Query Match      0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2732 gttgaacatttgag 2747
|||||
Db 229 GTTGAACATTGGAG 244

RESULT 73
US-08-461-360A-3
; Sequence 3, Application US/08461360A
; Patent No. 5650491
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,360A
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-461-360A-3
```

Query Match 0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2732 gttgaacattggag 2747
|||||
DB 229 GTTGAACATTGGAG 244

RESULT 74

US-08-461-359-3
; Sequence 3, Application US/08461359
; Patent No. 5686595
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,359
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9725
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; S-08-461-359-3

Query Match 0.5%; Score 16; DB 1; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2732 gttgaacattggag 2747
|||||
DB 229 GTTGAACATTGGAG 244

RESULT 75

PCT-US94-12904-3
; Sequence 3, Application PC/TUS9412904
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Takayama, Shinichi
; APPLICANT: Sato, Takaaki
; TITLE OF INVENTION: Bcl-2-associated proteins
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12904
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/152,485
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US94-12904-3

Query Match 0.5%; Score 16; DB 5; Length 733;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2732 gttgaacattggag 2747
|||||
DB 229 GTTGAACATTGGAG 244

Search completed: March 25, 2002, 18:01:41
Job time: 16640 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: March 25, 2002, 14:18:56 : Search time 270.41 Seconds

(without alignments)
9935.064 Million cell updates/sec

Title: US-09-697-089-1

Perfect score: 3133

Sequence: 1 cgcctagccgcggtgggaag.....aagtgactcgaagcagta 3133

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched:

Word size: 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : N.Geneseq_1101.*

- 1: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3133	100.0	3133	22	Human caspase 2
2	2969	94.8	3545	22	Human caspase 2
3	2634	84.1	3615	22	Human caspase 2
4	1807	57.7	2950	22	Human caspase 2
5	515	16.4	2735	22	Human caspase 2
6	242	7.7	421	22	Human caspase 2
7	242	7.7	421	22	Human caspase 2
8	242	7.7	421	22	Human caspase 2
9	220	7.0	220	22	Human caspase 2
10	220	7.0	220	22	Human caspase 2
11	220	7.0	220	22	Human caspase 2

12	19	0.6	549	22	AAH11452	Human CDNA clone (
13	19	0.6	579	19	AA330628	H. pylori cell env
14	19	0.6	649	19	AA144445	H. pylori GHPD 875
15	19	0.6	1908	21	AA13098	Aspergillus oryzae
16	19	0.6	2825	22	AAH16202	Human CDNA sequenc
17	18	0.6	116	21	AA14986	Human secreted pro
18	18	0.6	274	21	AAH31261	Plant microsatelli
19	18	0.6	322	16	AA124250	Human gene signatu
20	18	0.6	355	21	AAH31416	Plant microsatelli
21	18	0.6	377	21	AAH31325	Plant microsatelli
22	18	0.6	435	21	AAH31366	Plant microsatelli
23	18	0.6	446	21	AAH31366	Plant microsatelli
24	18	0.6	454	22	AAH14905	Human secreted pro
25	18	0.6	454	22	AAH14905	Human secreted pro
26	18	0.6	454	22	AAH14905	Human secreted pro
27	18	0.6	454	22	AAH14905	Human secreted pro
28	18	0.6	466	21	AAH14905	Human secreted pro
29	18	0.6	472	22	AAH14905	Human secreted pro
30	18	0.6	472	22	AAH14905	Human secreted pro
31	18	0.6	472	22	AAH14905	Human secreted pro
32	18	0.6	594	11	AAH05868	Sequence encoding
33	18	0.6	652	21	AAH07589	Fusarium venenatum
34	18	0.6	682	21	AAH04321	Arabidopsis thalia
35	18	0.6	814	22	AAH05406	Human CDNA clone (
36	18	0.6	888	20	AAH20212	Enterococcus faeca
37	18	0.6	960	22	AAH67086	C glutamicum codin
38	18	0.6	992	11	AAH05870	Sequence encoding
39	18	0.6	1034	21	AAH25257	Human secreted pro
40	18	0.6	1083	22	AAH71387	Corynebacterium gl
41	18	0.6	1205	20	AAH24236	Human normal blid
42	18	0.6	1250	9	AAH0316	Transcription cont
43	18	0.6	1473	21	AAH40012	Arabidopsis thalia
44	18	0.6	1491	21	AAH47150	DNA encoding a ser
45	18	0.6	1515	19	AAH40737	C. fells esterase,
46	18	0.6	1515	19	AAH40737	C. fells esterase,
47	18	0.6	1561	21	AAH21032	Human low adenosin
48	18	0.6	1561	21	AAH21032	Human low adenosin
49	18	0.6	1579	22	AAH13702	Human CDNA sequenc
50	18	0.6	1611	12	AAH12528	Thymidylate phosph
51	18	0.6	1611	21	AAH47151	DNA encoding a ser
52	18	0.6	1619	21	AAH15908	Human protein clon
53	18	0.6	1650	19	AAH40760	C. fells esterase,
54	18	0.6	1650	19	AAH40761	C. fells esterase,
55	18	0.6	1723	20	AAH40761	C. fells esterase,
56	18	0.6	1800	15	AAH67601	Human vesicular bi
57	18	0.6	1878	22	AAH25374	Retinoblastoma pro
58	18	0.6	1926	20	AAH82522	Nucleotide sequenc
59	18	0.6	1962	20	AAH77504	Splintomomas capsu
60	18	0.6	1982	19	AAH40735	C. fells esterase,
61	18	0.6	1982	19	AAH40735	C. fells esterase,
62	18	0.6	2144	19	AAH40758	C. fells esterase,
63	18	0.6	2144	19	AAH40758	C. fells esterase,
64	18	0.6	2299	19	AAH296326	S. pneumoniae derl
65	18	0.6	2299	19	AAH296326	S. pneumoniae derl
66	18	0.6	3000	20	AAH84399	Streptococcus pneu
67	18	0.6	3001	21	AAH51776	S. capsulata IPO12
68	18	0.6	3001	21	AAH51776	Chromosome 13q31-q
69	18	0.6	3205	21	AAH51787	Chromosome 13q31-q
70	18	0.6	3259	21	AAH25373	Genomic sequence o
71	18	0.6	3529	21	AAH31144	Human colon cancer
72	18	0.6	3942	20	AAH80604	Kidney injury asso
73	18	0.6	3969	21	AAH64473	DNA encoding an as
74	18	0.6	4215	22	AAH158415	Human polynucleoti
75	18	0.6	4871	22	AAH160201	Human polynucleoti
76	18	0.6	4951	20	AAH13198	Human poly nucleoti
77	18	0.6	4982	18	AAH85267	Enterococcus faeca
78	18	0.6	5212	9	AAH80317	HMC-CoA reductase
79	18	0.6	5292	21	AAH21374	Transcription cont
80	18	0.6	5511	17	AAH41853	Human low adenosin
81	18	0.6	5511	17	AAH41853	Human low adenosin
82	18	0.6	6212	21	AAH35253	Human low adenosin
83	18	0.6	7380	20	AAH84028	Human low adenosin
84	18	0.6	7383	22	AAH00153	Matrix metalloprot

c 85	18	0.6	7386	22	AA157671	Human colorectal c
c 86	18	0.6	45186	22	AA60478	Wld-type human CT
c 87	18	0.6	236303	22	AA511614	Human genomic DNA
c 88	18	0.6	349980	22	AA68530	C glutamincum codin
c 89	18	0.6	1664976	19	AAV21209	Methanococcus jann
c 90	17	0.5	21	20	AA218259	p450 enzyme gene s
c 91	17	0.5	47	21	AA267938	Human map-related
c 92	17	0.5	51	21	AA299916	Sequence of the st
c 93	17	0.5	121	21	AAAC10211	Human secreted pro
c 94	17	0.5	212	21	AAAC13747	Human secreted pro
c 95	17	0.5	276	20	AAV88937	EST clone HM846.
c 96	17	0.5	290	20	AA222458	Internal Transcrib
c 97	17	0.5	312	22	AAH66827	C glutamincum codin
c 98	17	0.5	333	21	AAAC22687	Human secreted exp
c 99	17	0.5	359	21	AAA44118	Human secreted exp
c 100	17	0.5	383	18	AA833360	Breast cancer tumo

ALIGNMENTS

SULF 1

AS03945

ID AAS03945 standard; CDNA: 3133 BP.

12-SEP-2001 (first entry)

Human caspase recruitment domain 12 (CARD-12) CDNA.

Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
 cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
 systemic lupus erythematosus; arthritis; neurological disorder; stroke;
 Alzheimer's disease; amyotrophic lateral sclerosis; haematologic disease;
 aplastic anaemia; myocardial infarction; inflammatory disorder;
 Crohn's disease; insulin-dependent diabetes; contact dermatitis;
 psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
 tuberculosis; ischaemic brain injury; hypoxic brain injury; ss;
 kidney ischaemia; reperfusion injury; acute bacterial meningitis;
 excitotoxic brain damage; liver disease.

Homo sapiens.

Key Location/Qualifiers
36..3110

/*tag= a /product= "Human CARD-12"

W0200130971-A2.

03-MAY-2001.

26-OCT-2000; 2000WO-US29643.

27-OCT-1999; 99US-0161822.

(MILL-) MILLENNIUM PHARM INC.

Berlin J, Robison KE;

WPI: 2001-308628/32.

P-PSDB: AAU02880.

Isolated caspase recruitment domain-12 polypeptide and nucleic acids

encoding them, useful for treating and diagnosing disorders associated

with abnormal apoptosis such as cancer, arthritis and Alzheimer's

disease - Claim 2: Fig 1: 93pp; English.

The sequence represents a CDNA which encodes the human caspase

recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a

number of proteins that transmit signals that activate apoptosis and
 inflammatory pathways in response to stress and other stimuli. Therefore,
 CARD-12 and its corresponding nucleic acid may be used in treatment and
 diagnosis of patients suffering from disorders associated with an
 abnormal level (an increase or a decrease) of apoptotic cell death or
 abnormal activity of stress-related pathways. The disorders include
 cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
 autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
 neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
 sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial
 infarction, stroke), inflammatory and immune system disorders (e.g.
 Crohn's disease, insulin-dependent diabetes, contact dermatitis,
 psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
 lepromatous leprosy), ischaemic and hypoxic brain injury, kidney
 ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
 meningitis and liver disease.

Sequence 3133 BP; 903 A; 691 C; 729 G; 810 T; 0 other;

Query Match 100.0%; Score 3133; DB 22; Length 3133;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	cgctctagccggttggaagcttccacagacaatgaattcataaggacatagcc	60
Db	1	cgctctagccggttggaagcttccacagacaatgaattcataaggacatagcc	60
Oy	61	gagcccttataaagaatggaatgactgtttataaacaatcacatgacctattg	120
Db	61	gagcccttataaagaatggaatgactgtttataaacaatcacatgacctattg	120
Oy	121	tatgaatgttcctgacgcgaagaagaatcatatttcgctcgaagaagtgagcag	180
Db	121	tatgaatgttcctgacgcgaagaagaatcatatttcgctcgaagaagtgagcag	180
Oy	181	atgcgtcgaagaggatcatcatgatttgaagaagggttcagagtcctgaacctc	240
Db	181	atgcgtcgaagaggatcatcatgatttgaagaagggttcagagtcctgaacctc	240
Oy	241	ttcttaaatcccttaagaagtggaacttccctctttccagacttgatgacaagtc	300
Db	241	ttcttaaatcccttaagaagtggaacttccctctttccagacttgatgacaagtc	300
Oy	301	ttttcatcagacatcagaagaagagcttgacagatttgctcagagattaaagactgt	360
Db	301	ttttcatcagacatcagaagaagagcttgacagatttgctcagagattaaagactgt	360
Oy	361	accatacccatctttctgaactttatccctctgtgtgaagatatgacattattta	420
Db	361	accatacccatctttctgaactttatccctctgtgtgaagatatgacattattta	420
Oy	421	acttgaagaacaccttcacagaacctgctcgttggaagaagaacacacacacacgcg	480
Db	421	acttgaagaacaccttcacagaacctgctcgttggaagaagaacacacacacacgcg	480
Oy	481	tggaagcagctgacaccttgaaatggcctcctcgcagagcttccatcagaagccctgcatatgag	540
Db	481	tggaagcagctgacaccttgaaatggcctcctcgcagagcttccatcagaagccctgcatatgag	540
Oy	541	gggaatctggcacaaggaagtcacactcgtcgcagagcattgacatgctctggggctcg	600
Db	541	gggaatctggcacaaggaagtcacactcgtcgcagagcattgacatgctctggggctcg	600
Oy	601	gaaatgcaagagcttcgcaagttcaaatcgtcttctccctcgtctcagcagggcc	660
Db	601	gaaatgcaagagcttcgcaagttcaaatcgtcttctccctcgtctcagcagggcc	660
Oy	661	agggtgagacttttgaacacctctgtgatacaactcctgataactcgtgacacatcagga	720
Db	661	agggtgagacttttgaacacctctgtgatacaactcctgataactcgtgacacatcagga	720
Oy	721	agcagacatcatgcatgctgctgaaagctgcgcagagaggttcttctcttgatg	780

[illegible]

Oy	2941	tttcttaacttagtactaaagaattcttactcagatccagaattagtcaagaacttgcc	3000
Db	2941	tcttcgtacttaagctaccaaagaattcttaccggtacccagcatltagtcaagaacttgcc	3000
Oy	3001	aagctgtatccaaagctttaacctttctcgcaagaagctaagcttgtgtggctgaattgatg	3060
Db	3001	aagctgtatccaaagctttaacctttctcgcaagaagctaggcttgtgtggctgaattgatg	3060
Oy	3061	atgatgatctcaagtgtgttatatacagagtgcttttaactcagtaactgccttaataaagtgtca	3120
Db	3061	atgatgatctcagctgtgtatatatacagagtgcttttaactcagtaactgccttaataaagtgtca	3120
Oy	3121	ctcgaagccaagta	3133
Db	3121	ctcgaagccaagta	3133
RESULT	2		
AAH98254			
AAH98254	standard; CDNA; 3545 BP.		
AAH98254;			
12-OCT-2001	(first entry)		
Murine EST-derived coding sequence SEQ ID NO: 111.			
Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;			
tomato; monkey; dog; sea urchin; expressed sequence tag; EST;			
diagnostics; forensic test; gene mapping; genetic disorder;			
biodiversity; gene therapy; nutrition; ss.			
Mus musculus.			
WO200154477-A2.			
02-AUG-2001.			
25-JAN-2001; 2001MO-USO2687.			
25-JAN-2000; 2000US-0491404.			
17-JUL-2000; 2000US-0617746.			
03-AUG-2000; 2000US-0631451.			
15-SEP-2000; 2000US-0663870.			
(HYSE-) HYSEQ INC.			
Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;			
Cao Y, Drmanac RA, Zhang J, Werhman T;			
WPI: 2001-476164/51.			
P-PSDB: AAM23595.			
Isolated polypeptide for treatment of diseases, diagnostics, raising			
antibodies and research use -			
Claim 1; Page 250-251; 1275pp; English.			
The present invention provides the protein and coding sequences of novel			
proteins from a variety of organisms, including human, dog, cat, horse,			
cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea			
urchin and tomato. These were derived from expressed sequence tags (ESTs)			
from the organism of interest. They can be used in diagnostics,			
forensics, gene mapping, identification of mutations, to assess			
biodiversity and for nutritional purposes. The present sequence is a cDNA			
of the invention.			
Sequence 3545 BP; 1038 A; 755 C; 816 G; 936 T; 0 other;			
Query Match	94.8%; Score 2969; DB 22; Length 3545;		
Best Local Similarity	99.9%; Pred No. 0;		

[illegible]

[illegible]

AC AAS03946;
XX 12-SEP-2001 (first entry)
DT Human caspase recruitment domain 12 (CARD-12) genomic DNA.
XX
DE
XX
KW Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KW Cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW Systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KW Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;
KW Aplastic anaemia; myocardial infarction; inflammatory disorder;
KW Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KW Psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW Tuberculosis; ischaemic brain injury; hypoxic brain injury; ds;
KW Kidney ischaemia; reperfusion injury; acute bacterial meningitis;
KW Excitotoxic brain damage; liver disease.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT 1..3615
CD CDS /tag_ a
/product= "Human CARD-12"
XX
XX MO200130971-A2.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 2000WO-US29643.
XX
XX 27-OCT-1999; 99US-0161822.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Bertin J, Robison KE;
XX
XX WPI: 2001-308628/32.
XX
XX P-PSDB; AAU02881.
XX
XX
XX Isolated caspase recruitment domain-12 polypeptide and nucleic acids
XX encoding them, useful for treating and diagnosing disorders associated
XX with abnormal apoptosis such as cancer, arthritis and Alzheimer's
XX disease -
XX
XX
XX Disclosure: Fig 2; 93pp; English.
XX
XX
XX The sequence represents a genomic DNA which encodes the human caspase
XX recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
XX number of proteins that transmit signals that activate apoptosis and
XX inflammatory pathways in response to stress and other stimuli. Therefore,
XX CARD-12 and its corresponding nucleic acid may be used in treatment and
XX diagnosis of patients suffering from disorders associated with an
XX abnormal level (an increase or a decrease) of apoptotic cell death or
XX abnormal activity of stress-related pathways. The disorders include
XX cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
XX autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
XX neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
XX sclerosis), hematologic diseases (e.g. aplastic anaemia, myocardial
XX infarction, stroke), inflammatory and immune system disorders (e.g.
XX Crohn's disease, insulin-dependent diabetes, contact dermatitis,
XX psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
XX lepromatous leprosy), ischaemic and hypoxic brain injury, kidney
XX ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial
XX meningitis and liver disease.
XX
XX
XX Sequence 3615 BP: 1041 A; 811 C; 845 G; 918 T; 0 other;

Query Match 84.1%; Score 2634; DB 22; Length 3615;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX 298 gcttttcatcagacatcagaagagacttgcacgattgtgctcagatttaagact 357

Db 734 gcttttcatcagacatcagaagagacttgcacgattgtgctcagatttaagact 793
QY 358 tgtaccatcccccatcttctcgaactttatccctctgtggaagatattgacattt 417
Db 794 tgtaccatcccccatcttctcgaactttatccctctgtggaagatattgacattt 853
QY 418 ttaacttgaagac 477
Db 854 ttaacttgaagac 913
QY 478 gctgtgagcagcttgacacacacacacacacacacacacacacacacacacac 537
Db 914 gctgtgagcagcttgacacacacacacacacacacacacacacacacacacac 973
QY 538 aaggggaactctgcaagagcagctcctcctcctcctcctcctcctcctcctcct 597
Db 974 aaggggaactctgcaagagcagctcctcctcctcctcctcctcctcctcctcct 1033
QY 598 ccggaagctgcaagagcagctcctcctcctcctcctcctcctcctcctcctcct 657
Db 1034 ccggaagagcagagcctcctcctcctcctcctcctcctcctcctcctcctcct 1093
QY 658 ccggaagctgcaagagcagctcctcctcctcctcctcctcctcctcctcctcct 717
Db 1094 ccggaagctgcaagagcagctcctcctcctcctcctcctcctcctcctcctcct 1153
QY 718 ggaagcagac 777
Db 1154 ggaagcagac 1213
QY 778 atggttacaatgaattcgaagcccgagacacacacacacacacacacacacac 837
Db 1214 atggttacaatgaattcgaagcccgagacacacacacacacacacacacacac 1273
QY 838 accacgccttcaagaacacacacacacacacacacacacacacacacacacac 897
Db 1274 accacgccttcaagaacacacacacacacacacacacacacacacacacacac 1333
QY 898 ggcagcttgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 957
Db 1334 ggcagcttgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1393
QY 958 tcatcgcgaagctgcttgaatcgaagcctgtgtgtgtgtgtgtgtgtgtgtgtgt 1017
Db 1394 tcatcgcgaagctgcttgaatcgaagcctgtgtgtgtgtgtgtgtgtgtgtgtgt 1453
QY 1018 ccaggttgcttgaggaatcctcgaagacacacacacacacacacacacacacac 1077
Db 1454 ccaggttgcttgaggaatcctcgaagacacacacacacacacacacacacacac 1513
QY 1078 agatgggtgaagatgagctcctcctcctcctcctcctcctcctcctcctcctcct 1137
Db 1514 agatgggtgaagatgagctcctcctcctcctcctcctcctcctcctcctcctcct 1573
QY 1138 atctgttatacagaacacacacacacacacacacacacacacacacacacac 1197
Db 1574 atctgttatacagaacacacacacacacacacacacacacacacacacacac 1633
QY 1198 ggaagcttgagcagcctgaggaacacacacacacacacacacacacacacacac 1257
Db 1634 ggaagcttgagcagcctgaggaacacacacacacacacacacacacacacacac 1693
QY 1258 tgaagctcagagatgtgccaagcgtgaatggaatgctcctcctcctcctcctcct 1317
Db 1694 tgaagctcagagatgtgccaagcgtgaatggaatgctcctcctcctcctcctcct 1753
QY 1318 gtaaatatacagctcacaagctcacaagcacaagcacaagcacaagcacaagcaca 1377
Db 1754 gtaaatatacagctcacaagctcacaagcacaagcacaagcacaagcacaagcaca 1813
QY 1378 agtacacagcagacagac 1437

Db	1814	agtacaacgacgagcaagaaagaccacacagtttatctgaagctctcatatgaagccagaagagtgta	1873
Qy	1438	ccaaggggaatggttactctgcagaaatggtttccatttcggaatatacatccactata	1497
Db	1874	ccaaggggaatggttactctgcagaaatggtttccatttcggaatatacatccactata	1933
Qy	1498	ggaagctctccggttaaacctctgtgtcatctctgtgaagccacagggctgtttagaagc	1557
Db	1934	ggaagctctccggttaaacctctgtgtcatctctgtgaagccacagggctgtttagaagc	1993
Qy	1558	acctcgacgacgtgatatacaacgagctgcctctcggaattccatccgaagggctc	1617
Db	1994	acctcgacgacgtgatatacaacgagctgcctctcggaattccatccgaagggctc	2053
Qy	1618	tcctgagacaggaatctctctgcaagtgctgaanaaacacacacttgagcaagaatcttgaaag	1677
Db	2054	tcctgagacaggaatctctctgcaagtgctgaanaaacacacacttgagcaagaatcttgaaag	2113
Qy	1678	ccataaatacatcaatctccttctgtagaagtgctgcatccattatataatgaagagttaccca	1737
Db	2114	ccataaatacatcaatctccttctgtagaagtgctgcatccattatataatgaagagttaccca	2173
Qy	1738	aatcgagccctgagccaagaatttgaaagcttcttcctcaaggttaaaagcttatatacaact	1797
Db	2174	aatcgagccctgagccaagaatttgaaagcttcttcctcaaggttaaaagcttatatacaact	2233
Qy	1798	cagggaacatcccgcatctactatctgaactctctcttgaaacatttgcccaattgtgcaagtg	1857
Db	2234	cagggaacatcccgcatctactatctgaactctctcttgaaacatttgcccaattgtgcaagtg	2293
Qy	1858	ctctggaacttaataactctggaactttatggggagagctatggtctcatatgaggaagagctg	1917
Db	2294	ccctggaacttaataactctggaactttatggggagagctatggtctcatatgaggaagagctg	2353
Qy	1918	cagaagacacaggtggaatccaacatgagaagggcccgcaaaacactacatcccgcaagtg	1977
Db	2354	cagaagacacaggtggaatccaacatgagaagggcccgcaaaacactacatcccgcaagtg	2413
Qy	1978	ctgtatctcttctctcaacttgaagcaggaattcagagactctggaaggttcaacatccggg	2037
Db	2414	ctgtatctcttctctcaacttgaagcaggaattcagagactctggaaggttcaacatccggg	2473
Qy	2038	attcagaagaatttgaataagcaagatatccatatctctggggaaaaatattcagctctgcca	2097
Db	2474	attcagaagaatttgaataagcaagatatccatatctctggggaaaaatattcagctctgcca	2533
Qy	2098	caagagctcagagctgtaaaataagaatagatgctcgtgtgtgagcctcggaagcctcagttgtgcc	2157
Db	2534	caagagctcagagctgtaaaataagaatagatgctcgtgtgtgagcctcggaagcctcagttgtgcc	2593
Qy	2158	tcagacactgttaagaacattatctctccatctgttgaagcagctccctccatccatagaag	2217
Db	2594	tcagacactgttaagaacattatctctccatctgttgaagcagctccctccatccatagaag	2653
Qy	2218	atgagaagcacatcacatctgttaacaaacctgtaaaaccttgaatttcaatgaactacaga	2277
Db	2654	atgagaagcacatcacatctgttaacaaacctgtaaaaccttgaatttcaatgaactacaga	2713
Qy	2278	atcaacgagctgcggggtgtctgcacttgacagctgtgtgtgaactttgaagaacctttcaagc	2337
Db	2714	atcaacgagctgcggggtgtctgcacttgacagctgtgtgtgaactttgaagaacctttcaagc	2773
Qy	2338	tcataatgtataaataagaatgtaatgaagaagatgctataaaactgaactgaagggccgga	2397
Db	2774	tcataatgtataaataagaatgtaatgaagaagatgctataaaactgaactgaagggccgga	2833
Qy	2398	aaaacctgaaagaatgctgtattatttcaatttgaccacctctgtgacatttggaaggggaa	2457
Db	2834	aaaacctgaaagaatgctgtattatttcaatttgaccacctctgtgacatttggaaggggaa	2893
Qy	2458	tgtgattatcatgtcaagctctctgtcaagtgaaaccttgagaccttgaagaatattcaatag	2517
Db	2894	tgtgattatcatgtcaagctctctgtcaagtgaaaccttgagaccttgaagaatattcaatag	2953

Qy	2518	ttccctgtctgtgtctgtcaaatgtcagtgtaaaatctctagctcagaatcttcacaatttgg	2577
Db	2954	ttccctgtctgtgtgtctgtcaaatgtcagtgtaaaatctctagctcagaatcttcacaatttgg	3013
Qy	2578	tcaaatgtagcatctcttgatttatccaagaataattcactctggaaaaaagatggaatgaagctc	2637
Db	3014	tcaaatgtagcatctcttgatttatccaagaataattcactctggaaaaaagatggaatgaagctc	3073
Qy	2638	ttcatgaactatgcagacaggaatgaacgtgctcagagaacagctcacacgacttgatctgcct	2697
Db	3074	ttcatgaactatgcagacaggaatgaacgtgctcagagaacagctcacacgacttgatctgcct	3133
Qy	2698	ggggctgtgacgtgcaagggcagccttgagcagctgttgaacatttggagaggtlccac	2757
Db	3134	ggggctgtgacgtgcaagggcagccttgagcagctgttgaacatttggagaggtlccac	3193
Qy	2758	aactcgtcaagctgtgggttgtaaaaaactcggagactcacagatacagagattagaatttag	2817
Db	3194	aactcgtcaagctgtgggttgtaaaaaactcggagactcacagatacagagattagaatttag	3253
Qy	2818	gtgcaatttttggaaagaacccctctgaaaaactccagcagctgaattttgctgggaatc	2877
Db	3254	gtgcaatttttggaaagaacccctctgaaaaactccagcagctgaattttgctgggaatc	3313
Qy	2878	gtgtgagcagtgatgagatgctgttcctctcatgggttgtaatttgaaatcttaagaacttag	2937
Db	3314	gtgtgagcagtgatgagatgctgttcctctcatgggttgtaatttgaaatcttaagaacttag	3373
Qy	2938	tgtttttgactttagtactaaagaattcttaacctgaatccagatcagatcagagaactta	2997
Db	3374	tgtttttgactttagtactaaagaattcttaacctgaatccagatcagatcagagaactta	3433
Qy	2998	gccaaagtataccaagttaactttctcgcaagaagcgtgagctgtgtgggtggcaatttg	3057
Db	3434	gccaaagtataccaagttaactttctcgcaagaagcgtgagctgtgtgggtggcaatttg	3493
Qy	3058	atgatgatcatcagtgattatcacg	3084
Db	3494	atgatgatcatcagtgattatcacg	3520

RESULT	4
AAH99581	
ID	AAH99581 standard; cDNA; 2950 BP.
XX	
AC	AAH99581;
XX	
DT	16-OCT-2001 (first entry)
XX	
DE	Human protein encoding cDNA sequence SEQ ID NO:416.
XX	
KW	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW	antiinflammatory; antirheumatic; antirheumatic; immunosuppressive;
KW	antibacterial; endocrine; cardiant; central nervous system; virucide;
KW	anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW	antiagregant; haemostatic; vulnerability; antilucer; osteopathic; eczema;
KW	dermatological; antiallergic; antiaesthetic; antidiabetic; cytosolic;
KW	neuroprotective; antidepressant; nootropic; antiparkinsonian; infection
KW	immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW	antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW	cardiac dysfunction; neurop

XX 22-DEC-2000; 2000WO-US35017.
XX
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX (HXSE-) HXSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-457603/49.
DR P-PSDB: AAM25640.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 1; Page 511-512; 1217pp; English.
XX
XX AAH9166 to AAH9904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virocidic; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianemic; antiagregant; haemostatic; vulnery;
CC antiulcer; osteopathic; dermatological; antiallergic; antidiabetic;
CC antidiabetic; cytosolic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neurophology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
XX Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other:
SO
Query Match 57.7%; Score 1807; DB 22; Length 2950;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1225 ctctggaggggtgtgttcccccagaagtttgatctggaactgagatgttgcacgctga 1284
b 45 ctctggaggggtgtgttcccccagaagtttgatctggaactgagatgttgcacgctga 104
QY 1285 atgagagatgtctctgacacactggctcctctgttaataatacaactcaaaagttcaagc 1344
Db 105 atgagagatgtctctgacacactggctcctctgttaataatacaactcaaaagttcaagc 164
QY 1345 caaagtataaattcttcaacagttcattccagagttacacagcaggaagaactcagca 1404
Db 165 caaagtataaattcttcaacagttcattccagagttacacagcaggaagaactcagca 224
QY 1405 gtttatatagcttcataagcagagaggtgacccaagggaatgttactctgacgaana 1464
Db 225 gtttatatagcttcataagcagagaggtgacccaagggaatgttactctgacgaana 284
QY 1465 tggttccatttcgacattacatccacttataagcagctctccggtacacactgtgagt 1524
Db 285 tggttccatttcgacattacatccacttataagcagctctccggtacacactgtgagt 344
QY 1525 catctgtggaagcacaagggctgttatgaagcaccttcgacagtgatatacaacggct 1584
Db 345 catctgtggaagcacaagggctgttatgaagcaccttcgacagtgatatacaacggct 404
QY 1585 gctcttcgagcttccatcgccaagagctctctctgagacagaactcttgcaaatg 1644

Db 405 gctcttcgagcttccatcgccaagagctctctctgagacagaactcttgcaaatg 464
QY 1645 tgaanaaacacacactgagcaagaattcttgaagaagcattaaacatcattctgttagagt 1704
Db 465 tgaanaaacacacactgagcaagaattcttgaagaagcattaaacatcattctgttagagt 524
QY 1705 gtgcatcatttataataagagagtagataccaaatcagccctgagccaagaatttgaag 1764
Db 525 gtgcatcatttataataagagagtagataccaaatcagccctgagccaagaatttgaag 584
QY 1765 cttcttccaagtagtaaaagcttataatacaactcaggggaacatcccccattattttg 1824
Db 585 cttcttccaagtagtaaaagcttataatacaactcaggggaacatcccccattattttg 644
QY 1825 acttcttgaacatttgcacatttgcgaatgtgcctggaacttcaattaaactgagcttt 1884
Db 645 acttcttgaacatttgcacatttgcgaatgtgcctggaacttcaattaaactgagcttt 704
QY 1885 atgaggagctatgtgcttcatttggaaaggtgcagaagacacaggtgaaatccacatg 1944
Db 705 atgaggagctatgtgcttcatttggaaaggtgcagaagacacaggtgaaatccacatg 764
QY 1945 aagagggcccaaaaacccaattcccaagcaggtctgtattcttcttcaacttgaagc 2004
Db 765 aagagggcccaaaaacccaattcccaagcaggtctgtattcttcttcaacttgaagc 824
QY 2005 aggaattcaagactcttgaggtgcacacccagatcttaagaagaatttgaataagaagata 2064
Db 825 aggaattcaagactcttgaggtgcacacccagatcttaagaagaatttgaataagaagata 884
QY 2065 tcaatattctgggaaataatattcaagctctgcacaagcctcaagctgcaataaagagat 2124
Db 885 tcaatattctgggaaataatattcaagctctgcacaagcctcaagctgcaataaagagat 944
QY 2125 gtgcgtgtgtgtgtgaaagccttctgttctgaagacccgtgaagaacttattcttc 2184
Db 945 gtgcgtgtgtgtgtgaaagccttctgttctgaagacccgtgaagaacttattcttc 1004
QY 2185 tcatgtgtgaaagccagctccctccacacataagaagatgaaagacatccatctgtaacaa 2244
Db 1005 tcatgtgtgaaagccagctccctccacacataagaagatgaaagacatccatctgtaacaa 1064
QY 2245 acctgaaacacttgatattcaatgaactcaagaatcaacagcgtgcgggtgtctgactg 2304
Db 1065 acctgaaacacttgatattcaatgaactcaagaatcaacagcgtgcgggtgtctgactg 1124
QY 2305 acagcttgggttaacttgaagaaccttacaagaactatagatgatacgaataaagatgaatg 2364
Db 1125 acagcttgggttaacttgaagaaccttacaagaactatagatgatacgaataaagatgaatg 1184
QY 2365 aagaagatgctataaatactgaagccttgaagaaccttgaagaaccttgaagaaccttgaag 2424
Db 1185 aagaagatgctataaatactgaagccttgaagaaccttgaagaaccttgaagaaccttgaag 1244
QY 2425 atttgacccactgtctgacatttgaagaggaatggaattacatagctcactgtctca 2484
Db 1245 atttgacccactgtctgacatttgaagaggaatggaattacatagctcactgtctca 1304
QY 2485 gtgaacctgtgaccttgaagaataatcaattagctctctgctgtgtgtcgaatgag 2544
Db 1305 gtgaacctgtgaccttgaagaataatcaattagctctctgctgtgtgtcgaatgag 1364
QY 2545 tgaataatcctagctcagatcttcaaatgttgcataaactgagcaattcttattatcag 2604
Db 1365 tgaataatcctagctcagatcttcaaatgttgcataaactgagcaattcttattatcag 1424
QY 2605 aaatctaccttgaaagaatggaagaatggaagctctcatgaaactgatacagaagatgaag 2664
Db 1425 aaatctaccttgaaagaatggaagaatggaagctctcatgaaactgatacagaagatgaag 1484
QY 2665 tgcctagaacagctacacagctatgtgctgccttgagggtgtgtgaagtgaaagcagcctga 2724

DB 1485 tgcataaacagctcacagcactgatgctccctgggctgtgacgtcgaagcagccctga 1544
QY 2725 gcaagctgttgaacaactttggaggaagtcaccaactcgtcaagcttgggttgaanaact 2784
DB 1545 gcaagctgttgaacaactttggaggaagtcaccaactcgtcaagcttgggttgaanaact 1604
QY 2785 ggaagctcacagatacagagattagattttagtgctattttttggaagaacccctcga 2844
DB 1605 ggaagctcacagatacagagattagattttagtgctattttttggaagaacccctcga 1664
QY 2845 aaaaacttcagcagtgtaatttggcggaagtcgtgacagtgatgagtgagctgcct 2904
DB 1665 aaaaacttcagcagtgtaatttggcggaagtcgtgacagtgatgagtgagctgcct 1724
QY 2905 tcatgggtgtatttgaagatcttaagcaattagtgcttttttgaactttagtactcaagaat 2964
DB 1725 tcatgggtgtatttgaagatcttaagcaattagtgcttttttgaactttagtactcaagaat 1784
Y 2965 tctaccctgatacagatagtcagagaacttagcgaagtgatgatacgaagttacttcc 3024
DB 1785 tctaccctgatacagatagtcagagaacttagcgaagtgatgatacgaagttacttcc 1844
QY 3025 tgcagaagcagtagagctgtgtgggtggcgaatttgatgatactcagtgatatacag 3084
DB 1845 tgcagaagcagtagagctgtgtgggtggcgaatttgatgatactcagtgatatacag 1904
QY 3085 gtgcttttaactagtaactgtcttaataagtgtactcgaagccagta 3133
DB 1905 gtgcttttaactagtaactgtcttaataagtgtactcgaagccagta 1953

RESULT 5
AAH34171/c
ID AAH34171 standard; cDNA: 2735 BP.
AC AAH34171;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1253.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
OS Homo sapiens.
XX
NN WO200122920-A2.
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
XX
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
XX
DR P-PSDB: AAG74766.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3017; 9803BP; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37296 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 2735 BP; 791 A; 555 C; 512 G; 876 T; 1 other;

Query Match 16.4%; Score 515; DB 22; Length 2735;
Best Local Similarity 99.8%; Pred. No. 2,1e-245;
Matches 565; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2568 cacaatttggccaactgagcattcttgatattacagaanaattacctggaaaagatgga 2627
DB 2727 CACAATTGGTCAAACTGAGCATTCTTGATTATCAGAAAATTACCTGGAAGAAATGGA 2668
QY 2628 aatgaagctcttcatactgatacagacagatgaacgtgtctaaacagctcacgcagctg 2687
DB 2667 AATGAAGCTCTTCATGATGACGATGACAGGATGACGCTGTAACAGCTCACCGCAGCTG 2608
QY 2688 atgctgccctggggcgctgtgacgtgcaagcagcctgagcagctgttgaacaatttggag 2747
DB 2607 ATGCTGCCCTGGGGCGCTGTGACGTGCAAGGACGCTGAGCAGCTGTGTTGAACATTTGGAG 2548
QY 2748 gaagtcaccacaactcgtlcaagcttgggttgaanaacttggagactcagatatacagatt 2807
DB 2547 GAGGTCCACAACTCGTCAAGCTTGGGTGAAAACCTGAGACTCACAGATACAGAGATT 2488
QY 2808 agaattttagtgatcttttttggaaaagaaacccctgaanaacttccagagtgtaattg 2867
DB 2487 AGAATTTTAGTGCATTTTGGAAAAGAACCCCTGTGAAAACCTTCAGAGCTTGAATTTG 2428
QY 2868 ggcggaatcgtgtgacagtgatgagtgagctgtgcttcatacagtggttatttgaagactc 2927
DB 2427 GCGGGAATCGTGTGAGCAGTATGATGATGCTTGCCCTCATGGGTGATTTGGAATCTT 2368
QY 2928 aagcaattagtgcttttttgaacttgaactaaagaatttctacatccagatcagctagtc 2987
DB 2367 AAGCAATTAGTCTTTTGTGACTTACTTAAGAAATTTCTACCTGATCCAGCATTAGTC 2308
QY 2988 agaaaacttagcgaagtgatatacgaagttacttctcgaagaacctaagctgtgtgg 3047
DB 2307 AGAAAACCTTAGCCAGAGTGTATCCAGATTACTTTCTGCAAGACCTAGCCTTGTGGG 2248
QY 3048 tggcaatttgatgatgatgatacagtgatgatacagtgatgatacagtgatgatacagtgat 3107
DB 2247 TGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2188
QY 3108 taaataaagtgtactcgaagccagta 3133
DB 2187 TAAATAAAGTGTACTCGAAGCCAGTA 2162

RESULT 6
AAH14389
ID AAH14389 standard; DNA: 421 BP.
XX
XX AAH14389;
AC
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #4322 for gene expression analysis in human cervical cell sample.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;

```
KW cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID NO 4322; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;
XX
XX
XX Query Match 7.7%; Score 242; DB 22; Length 421;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-110;
XX Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2816 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 2875
XX |||||||
XX DB 180 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 239
XX |||||||
XX Y 2876 tcgtgtgagcagtgatgagtgctgccttcattgggtgtattggaactcttaagaact 2935
XX |||||||
XX DB 240 tcgtgtgagcagtgatgagtgctgccttcattgggtgtattggaactcttaagaact 299
XX |||||||
XX QY 2936 agtgttttttgaacttgaactaagaattctctacatgacagatagtcagaact 2995
XX |||||||
XX DB 300 agtgttttttgaacttgaactaagaattctctacatgacagatagtcagaact 359
XX |||||||
XX QY 2996 tagccaagtgtatccaagttaacttctcgaagaactagagctgtgtgggtgcaatt 3055
XX |||||||
XX DB 360 tagccaagtgtatccaagttaacttctcgaagaactagagctgtgtgggtgcaatt 419
XX |||||||
XX QY 3056 tg 3057
XX ||
XX DB 420 tg 421
XX
XX RESULT 7
XX AA135764
XX ID AA135764 standard; DNA; 421 BP.
XX
XX AC, AA135764;
```

```
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #4450 used to measure gene expression in human placenta sample.
XX
XX DE Probe; microarray; human; placenta; antenatal diagnosis;
XX
XX KW genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 4450; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;
XX
XX
XX Query Match 7.7%; Score 242; DB 22; Length 421;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-110;
XX Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2816 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 2875
XX |||||||
XX DB 180 aggtgcatcttttggaaagaccctctgaaacctccagcagttgaattggcgggaaa 239
XX |||||||
XX QY 2876 tcgtgtgagcagtgatgagtgctgccttcattgggtgtattggaactcttaagaact 2935
XX |||||||
XX DB 240 tcgtgtgagcagtgatgagtgctgccttcattgggtgtattggaactcttaagaact 299
XX |||||||
XX QY 2936 agtgttttttgaacttgaactaagaattctctacatgacagatagtcagaact 2995
XX |||||||
XX DB 300 agtgttttttgaacttgaactaagaattctctacatgacagatagtcagaact 359
XX |||||||
XX QY 2996 tagccaagtgtatccaagttaacttctcgaagaactagagctgtgtgggtgcaatt 3055
XX |||||||
XX DB 360 tagccaagtgtatccaagttaacttctcgaagaactagagctgtgtgggtgcaatt 419
XX |||||||
XX QY 3056 tg 3057
XX ||
XX DB 420 tg 421
XX
XX RESULT 8
XX AA104213
XX ID AA104213 standard; DNA; 421 BP.
XX
XX
```

AC AAI04213;
XX
XX 09-OCT-2001 (first entry)
DT
XX
DE Probe #4204 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
KM Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX
XX WO200157270-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 29-JAN-2001; 2001WO-US00661.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-476286/51.
DR
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
PT
XX
XX Claim 25; SEQ ID NO 4204; 322pp; English.
PS
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;
SQ

Query Match 7.7%; Score 242; DB 22; Length 421;
Best Local Similarity 100.0%; Pred. No. 7.6e-110;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2816 aggtgcatttttggaaagaacctctgaaataattccagcagttgaatttggcgggaaa 2875
DB 180 aggtgcatttttggaaagaacctctgaaataattccagcagttgaatttggcgggaaa 239
QY 2876 tctgtgaagcagtgatgagcttgcctcatggtgataattgaagatttaagaatt 2935
DB 240 tctgtgaagcagtgatgagcttgcctcatggtgataattgaagatttaagaatt 299
QY 2936 agtgcatttttgaacttgaactaaagaatttctaccgtaccagcagcattagtcagaaact 2995
DB 300 agtgcatttttgaacttgaactaaagaatttctaccgtaccagcagcattagtcagaaact 359
QY 2996 tagcagaagtattacagtaactttctgcagaagaagcagcttcttgggtggcaat 3055
DB 360 tagcagaagtattacagtaactttctgcagaagaagcagcttcttgggtggcaat 419

QY 3056 tg 3057
DB 420 tg 421

RESULT 9
ID AAI23590
AAI23590 standard; DNA: 220 BP.
AC
XX
XX AAI23590;
DT
XX
XX 12-OCT-2001 (first entry)
DE
XX
XX Probe #13523 for gene expression analysis in human cervical cell sample.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KM cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157278-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488901/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PT
XX
XX Claim 25; SEQ ID NO 13523; 487pp; English.
PS
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;
SQ

Query Match 7.0%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2865 ttgcgggaaatcggtgagcagtgatgagcttgcctcatggtgataattgagaat 2924
DB 1 ttgcgggaaatcggtgagcagtgatgagcttgcctcatggtgataattgagaat 60
QY 2925 cttaagaattaggttttgaacttaagtaactaaagaatttctaccgtaccagcacta 2984
DB 61 cttaagaattaggttttgaacttaagtaactaaagaatttctaccgtaccagcacta 120
QY 2985 gtccagaacttagccagtgattacccaagttaactttctgcagaagaagcagcttctt 3044

Db 121 gtcagaaaacttagccaaagtgatcatccaaagtaactttctgcagaagctagcttggt 180

Qy 3045 gggctggcaattgtagatgatgatctcaagtgattatacag 3084
|||||
Db 181 gggctggcaattgtagatgatgatctcaagtgattatacag 220

RESULT 10
AA148904
ID AA148904 standard; DNA; 220 BP.
AC AA148904;
XX
XX 17-OCT-2001 (first entry)
DE Probe #17590 used to measure gene expression in human placenta sample.
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX Homo sapiens.
OS
LN WO200157272-A2.
IN
XX
PD 09-AUG-2001.
XX
PD 30-JAN-2001; 2001WO-US000663.
XX
PF 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WP1: 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 17590; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match 7.0%; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.2e-99;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2865 ttggggggaatcgtgtgtagcagtgatgtaggcttgcttcatactggtgatttgaagaat 2924
|||||
Db 1 ttggggggaatcgtgtgtagcagtgatgtaggcttgcttcatactggtgatttgaagaat 60

Qy 2925 cttaagcaattagtggttttttgaacttagtaactaaagaattctacccgatccagcatca 2984
|||||
Db 61 cttaagcaattagtggttttttgaacttagtaactaaagaattctacccgatccagcatca 120

Qy 2985 gtccagaaaacttagcacaagtgtatccaaagttaactttcttcgcaagaactagagcttggt 3044
|||||
Db 121 gtccagaaaacttagcacaagtgtatccaaagttaactttcttcgcaagaactagagcttggt 180

Qy. 3045 ggtgtggcaattgtagatgatgatctcaagtgattatacag 3084

[illegible]

OY 2985 gtcagaaacttagccaagtgtatccaagttctctgcaagaagctagctgtc 3044
 |||||||
 Db 121 gtcagaaacttagccaagtggtatccaagttctctgcaagaagctagctgtc 180
 |||||||
 OY 3045 gggctgcaattgcatgcatgcatcagtcgtattacag 3084
 |||||||
 Db 181 gggctgcaattgcatgcatgcatcagtcgtattacag 220
 |||||||

RESULT 12
 AAH1452
 ID AAH1452 standard; cDNA; 549 BP.
 AC AAH1452;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer) SEQ ID NO: 8287.
 XX
 KM Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 3; SEQ ID 8287; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH16332
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 549 BP; 162 A; 120 C; 119 G; 139 T; 9 other;

Query Match 0.6%; Score 19; DB 22; Length 549;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 caagaatgggaatgactg 90
 |||||||
 Db 214 caagaatgggaatgactg 232

RESULT 13
 AAX30628/c
 ID AAX30628 standard; DNA; 579 BP.
 AC AAX30628;
 XX
 DT 08-JUN-1999 (first entry)
 XX
 DE H. pylori cell envelope protein ORF hp7el0433_5345837_c2-8.
 XX
 KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KW secreted protein; cytoplasmic protein; cellular protein; ds.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9824475-A1.
 XX
 PD 11-JUN-1998.
 XX
 PF 05-DEC-1997; 97WO-US22104.
 XX
 PR 14-JUL-1997; 97US-0891928.
 PR 05-DEC-1996; 96US-0759625.
 PR 25-MAR-1997; 97US-0823745.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Alm RA, Castriotta LM, Dolg PC, Kabok Z, Smith D;
 XX
 DR WPI; 1998-333051/29.
 DR P-PSDB; AAY11099.
 XX
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 XX
 PS Claims 3, 4; Page 169; 339pp; English.
 XX
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC sequences containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 XX
 SQ Sequence 579 BP; 173 A; 130 C; 98 G; 178 T; 0 other;

Query Match 0.6%; Score 19; DB 19; Length 579;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 409 acattattttacttga 427
 |||||||
 Db 356 ACATATTTTTACTTGA 338

RESULT 14
AAI14445/c
ID AAI14445 standard; DNA: 649 BP.
XX
AC AAI14445;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHP0 875 gene.
XX
KM GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
KM peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 19..624
FT /tag= a
XX
PN MO9843478-A1.
XX
DR 08-OCT-1998.
XX
PF 01-APR-1998; 98MO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (IMMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI: 1998-542293/46.
DR P-PSDB; AAM98726.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 1538; 2054pp; English.
XX
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHP0 protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 649 BP; 191 A; 139 C; 118 G; 201 T; 0 other;

Query Match 0.6%; Score 19; DB 19; Length 649;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 409 acattatttacttgaa 427
|||||
DB 401 ACATTATTTTACTTGAA 383

RESULT 15
AAFI3098
ID AAFI3098 standard; cDNA: 1908 BP.
XX
AC AAFI3098;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:5621.

XX
KM Multiple gene expression; filamentous fungal cell; EST;
KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KM culture condition; environmental stress; spore morphogenesis;
KM metabolic pathway engineering; catalytic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000MO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI: 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2332; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catalytic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 1908 BP; 441 A; 497 C; 466 G; 496 T; 8 other;

Query Match 0.6%; Score 19; DB 21; Length 1908;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 555 ggcagtcacactgtctgc 573
|||||
DB 883 ggcagtcacactgtctgc 901

RESULT 16
AAHI6202/c
ID AAHI6202 standard; cDNA: 2825 BP.
XX
AC AAHI6202;
XX

XX WO9967421-A1.
 XX 29-DEC-1999.
 XX 25-JUN-1999; 99WO-N200092.
 XX 25-JUN-1998; 98US-0105307.
 XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX Havukkala IJ, Bloksberg LN, Glenn M;
 XX WPI; 2000-116958/10.
 XX New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 PS Claim 1; Page 140-141; 392pp; English.
 XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 SQ Sequence 274 BP; 59 A; 79 C; 75 G; 60 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 274;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatgatcgcag 3073
 |||||||
 DB 101 TGATGATGATGATCTCAG 84

RESULT 19
 AAT24250/C
 ID AAT24250 standard; cDNA to mRNA; 322 BP.
 XX AAT24250;
 XX 16-OCT-1996 (first entry)
 XX Human gene signature HUMGS06268.
 DE Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX Homo sapiens.
 OS WO9514772-A1.
 XX 01-JUN-1995.
 XX 11-NOV-1994; 94WO-JP01916.
 XX 12-NOV-1993; 93JP-0355504.
 XX (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX

PI Matsubara K, Okubo K;
 XX WPI; 1995-206931/27.
 DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
 XX for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX Claim 1; Page 1566; 2245pp; Japanese.
 XX A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 SQ Sequence 322 BP; 84 A; 53 C; 55 G; 119 T; 11 other;

Query Match 0.6%; Score 18; DB 16; Length 322;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 atgaatttcataaagac 53
 |||||||
 DB 51 ATGAATTCATTAAGGAC 34

RESULT 20
 AAA31416/C
 ID AAA31416 standard; DNA; 355 BP.
 XX AAA31416;
 XX 05-JUL-2000 (first entry)
 XX Plant microsatellite marker #377.
 DE Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.
 XX Eucalyptus grandis.
 OS WO9967421-A1.
 XX 29-DEC-1999.
 XX 25-JUN-1999; 99WO-N200092.
 XX 25-JUN-1998; 98US-0105307.
 XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX Havukkala IJ, Bloksberg LN, Glenn M;
 XX WPI; 2000-116958/10.
 XX New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -

PS Claim 1; Page 188; 392pp; English.

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences and associated flanking species. The sequences comprise a central core repeat sequence, especially selected from the sequences AAA32094-A32096 with left and right flanking sequences. The polynucleotide sequences can be used in the detection of DNA polymorphisms, in genome mapping, in physical mapping, in positional cloning of genes, in variety identification and in evaluation of genetic variability within and between plant tissues, populations, cultivars, species and species groups. They may also be used to design hybridization probes for oligonucleotide fingerprinting and library screening and to design primers for microsatellite-primed PCR. Microsatellite markers are useful to locate specific economically useful genes in plant genomes.

SQ Sequence 355 BP; 69 A; 98 C; 119 G; 68 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatctcag 3073
|||||
DB 101 TGATGATGATGATCTCAG 84

RESULT 21

AAA31325/C
ID AAA31325 standard; DNA; 377 BP.

AC AAA31325;
XX

DT 05-JUL-2000 (first entry)

DE Plant microsatellite marker #286.

XX

KW Plant microsatellite sequence; core repeat sequence; detection; probe; DNA polymorphism; genome mapping; physical mapping; fingerprinting; variety identification; genetic variability evaluation; primer; ss.

OS Eucalyptus grandis.

PN WO967421-A1.

PD 29-DEC-1999.

XX

25-JUN-1999; 99WO-N200092.

XX

25-JUN-1998; 98US-0105307.

PR

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Havukkala JU, Bloksberg LN, Glenn M;

XX

WPI: 2000-116958/10.

DR

XX

PT New plant microsatellite markers and associated flanking species for the detection of polymorphic genetic markers -

PT

PS Claim 1; Page 161; 392pp; English.

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences and associated flanking species. The sequences comprise a central core repeat sequence, especially selected from the sequences AAA32094-A32096 with left and right flanking sequences. The polynucleotide sequences can be used in the detection of DNA polymorphisms, in genome mapping, in physical mapping, in positional cloning of genes, in variety identification and in evaluation of genetic variability within and between plant tissues, populations, cultivars, species and species groups. They may also be used to design hybridization probes for oligonucleotide fingerprinting and library screening and to design

CC primers for microsatellite-primed PCR. Microsatellite markers are useful to locate specific economically useful genes in plant genomes.

XX

SQ Sequence 377 BP; 72 A; 102 C; 130 G; 73 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatctcag 3073
|||||
DB 88 TGATGATGATGATCTCAG 71

RESULT 22

AAA31366/C
ID AAA31366 standard; DNA; 435 BP.

AC AAA31366;
XX

DT 05-JUL-2000 (first entry)

DE Plant microsatellite marker #327.

XX

KW Plant microsatellite sequence; core repeat sequence; detection; probe; DNA polymorphism; genome mapping; physical mapping; fingerprinting; variety identification; genetic variability evaluation; primer; ss.

OS Eucalyptus grandis.

PN WO967421-A1.

PD 29-DEC-1999.

XX

25-JUN-1999; 99WO-N200092.

XX

25-JUN-1998; 98US-0105307.

PR

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
(FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Havukkala JU, Bloksberg LN, Glenn M;

XX

WPI: 2000-116958/10.

DR

XX

PT New plant microsatellite markers and associated flanking species for the detection of polymorphic genetic markers -

PT

PS Claim 1; Page 173; 392pp; English.

CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences and associated flanking species. The sequences comprise a central core repeat sequence, especially selected from the sequences AAA32094-A32096 with left and right flanking sequences. The polynucleotide sequences can be used in the detection of DNA polymorphisms, in genome mapping, in physical mapping, in positional cloning of genes, in variety identification and in evaluation of genetic variability within and between plant tissues, populations, cultivars, species and species groups. They may also be used to design hybridization probes for oligonucleotide fingerprinting and library screening and to design primers for microsatellite-primed PCR. Microsatellite markers are useful to locate specific economically useful genes in plant genomes.

CC

SQ Sequence 435 BP; 85 A; 125 C; 146 G; 79 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatctcag 3073
|||||

Db 101 TGATGATGATCTCAG 84

RESULT 23
AAC09333/C
ID AAC09333 standard; cDNA: 446 BP.
XX
AC AAC09333;
XX
06-OCT-2000 (first entry)
XX
Human secreted protein 5' EST, SEQ ID NO: 13408.
DE
XX
Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 9905-0122487.
XX
PA (GSEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 13408; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
XX Sequence 446 BP; 134 A; 76 C; 70 G; 162 T; 4 other;

Query Match 0.6%; Score 18; DB 21; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 atgaattcataagac 53
IIIIIIIIIIIIIIIIIIII
Db 106 ATGAATTCATTAAGAC 89

RESULT 24
AA114905/C
ID AA114905 standard; DNA: 454 BP.
XX
AC AA114905;
XX
XX 12-OCT-2001 (first entry)
XX

DE Probe #4838 for gene expression analysis in human cervical cell sample.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-488901/53.
XX
XX Claim 25; SEQ ID No 4838; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 ggggaatctggcaagac 557
IIIIIIIIIIIIIIIIIIII
Db 29 GGGGAATCTGGCAAGAC 12

RESULT 25
AA136257/C
ID AA136257 standard; DNA: 454 BP.
XX
AC AA136257;
XX
XX 17-OCT-2001 (first entry)
XX
DE Probe #4943 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
XX 09-AUG-2001.
XX

```
XX 30-JAN-2001; 2001WO-US00663.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 4943; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 18; DB 22; Length 454;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 540 ggggaatctggcaagc 557
DB 29 GGGGAATCTGCAAGGC 12
|||||
|||||

RESULT 26
AA104685/C
ID AA104685 standard; DNA; 454 BP.
XX
AC AA104685;
XX
XX 09-OCT-2001 (first entry)
XX
DE Probe #4676 used to measure gene expression in human breast sample.
XX
KM Probe: human; breast disease; breast cancer; development disorder; ss;
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
```

```
DR WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 4676; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 454 BP; 119 A; 89 C; 110 G; 136 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 18; DB 22; Length 454;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 540 ggggaatctggcaagc 557
DB 29 GGGGAATCTGCAAGGC 12
|||||
|||||

RESULT 27
AAC01809
ID AAC01809 standard; CDNA; 466 BP.
XX
AC AAC01809;
XX
XX 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1807.
XX
XX Human: 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR P-PSDB; AAC01803.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 1807; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
```

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

XX Sequence 466 BP; 76 A; 106 C; 132 G; 148 T; 4 other;

Query Match 0.6%; Score 18; DB 21; Length 466;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 766 ttctcctctgtagcgt 763
 ||||||||||||||||
 DB 220 ttctcctctgtagcgt 237

RESULT 28

AAA31287/C
 ID AAA31287 standard; DNA; 468 BP.

XX AAA31287;

DT 05-JUL-2000 (first entry)

XX Plant microsatellite marker #248.

XX Plant microsatellite sequence; core repeat sequence; detection; probe;
 KM DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KM variety identification; genetic variability evaluation; primer; ss.
 OS Eucalyptus grandis.

PN W09967421-A1.

XX 29-DEC-1999.

PF 25-JUN-1999; 99WC-NZ00092.

XX 25-JUN-1998; 98US-0105307.

XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Havukkala II, Bloksberg LN, Glenn M;

DR WPI; 2000-116958/10.

XX New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 PS Claim 1; Page 149; 392pp; English.

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.

XX Sequence 468 BP; 95 A; 129 C; 156 G; 87 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 468;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3056 tgatgatgatgattcag 3073
 ||||||||||||||||
 DB 101 TGATGATGATGATCAG 84

RESULT 29

AA11787
 ID AA11787 standard; DNA; 472 BP.

XX AA11787;

DT 12-OCT-2001 (first entry)

XX Probe #1720 for gene expression analysis in human cervical cell sample.

XX Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.

XX Homo sapiens.

OS W0200157278-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 1720; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 agaagtgagcaagatc 183
 ||||||||||||||||
 DB 143 agaagtgagcaagatc 160

```
RESULT 30
AAI33101
ID AAI33101 standard; DNA; 472 BP.
XX
AC AAI33101;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #1787 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KM genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 1787; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
Q Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;
XX
Query Match 0.6%; Score 18; DB 22; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 agaaggtgagcagatg 183
Db 143 agaaggtgagcagatg 160
IIIIIIIIIIIIIIIIIIII
RESULT 31
AAI01718
ID AAI01718 standard; DNA; 472 BP.
XX
AC AAI01718;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #1709 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast disease; breast cancer; development disorder; ss;
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
```

```
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
PS Claim 25; SEQ ID No 1709; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 472 BP; 129 A; 77 C; 153 G; 113 T; 0 other;
XX
Query Match 0.6%; Score 18; DB 22; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 agaaggtgagcagatg 183
Db 143 agaaggtgagcagatg 160
IIIIIIIIIIIIIIIIIIII
RESULT 32
AAQ05868/C
ID AAQ05868 standard; DNA; 594 BP.
XX
AC AAQ05868;
XX
DT 07-JAN-1991 (first entry)
XX
DE Sequence encoding mammalian growth hormone receptor binding protein.
XX
KW Ovine placental lactogen; ds;
XX
OS Ovis ammon aries.
XX
PN EP386979-A.
XX
PD 12-SEP-1990.
XX
PF 05-MAR-1990; 90EP-0302322.
XX
PR 06-MAR-1989; 89US-0319585.
```

XX (GERTH) GENENTECH INC.
 XX
 XX Wood WI, Colosi PC;
 XX
 DR WPI: 1990-276995/37.
 DR P-PSDB; AAR06643.
 XX
 PT DNA sequence encoding growth-hormone-receptor-binding protein -
 PT useful for increase of mammalian meat and milk prodn. and
 PT decrease in fat content.
 XX
 PS Claim 5; Page 9; 22pp; English.
 XX
 CC Receptor binding protein may be used to increase meat and milk
 CC production and decrease fat content especially in sheep and goats.
 CC Sequence shows an identity of 26% with human growth hormone and 67%
 CC to ovine placental lactogen.
 CC
 SO Sequence 594 BP; 191 A; 150 C; 125 G; 128 T; 0 other;

Query Match 0.6%; Score 18; DB 11; Length 594;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 535 ttgaagggaactcgaca 552
 ||||||||||||||||
 Db 54 TTGAAGGGGCAATCTGGCA 37

RESULT 33
 AAF07589
 ID AAF07589 standard; cDNA; 652 BP.
 XX
 AC AAF07589;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:112.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI: 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 419; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal

CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

Sequence 652 BP; 142 A; 185 C; 201 G; 121 T; 3 other;

Query Match 0.6%; Score 18; DB 21; Length 652;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2127 gctggtgctgctgaagc 2144
 ||||||||||||||||
 Db 578 gctggtgctgctgaagc 595

RESULT 34
 AAC44321/c
 ID AAC44321 standard; DNA; 682 BP.
 XX
 AC AAC44321;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 42431.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126284.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132488.
PR 07-MAY-1999; 9905-0132489.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135529.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 14-JUN-1999; 9905-0138847.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144684.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.

PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 16-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-01582029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.6%; Score 18; DB 21; Length 682;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

463 accacaccatcacgcg 480
|||||
517 ACCAACACCATCACCGC 500

DB 517 ACCAACACCATCACCGC 500

RESULT 35
AAH05406/c
ID AAH05406 standard; cDNA: 814 BP.
AC AAH05406;
DT 26-JUN-2001 (first entry)
DE Human cDNA clone (5'-primer) SEQ ID NO:2241.
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
OS Homo sapiens.
PN EP1074617-A2.
PD 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX WPI: 2001-318749/34.
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 1; SEQ ID 2241; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SQ Sequence 814 BP; 155 A; 295 C; 222 G; 136 T; 6 other;

Query Match 0.6%; Score 18; DB 22; Length 814;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 904 ttgtgacctgactgctg 921
|||||
DB 786 TTGTGTCCCTGACTGCTG 769

RESULT 36
AAAX20212
ID AAAX20212 standard; DNA: 888 BP.
AC AAAX20212;
DT 20-APR-1999 (first entry)
DE Enterococcus faecalis gene EF110.
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX detection; attenuation; antigenic; ss.
OS Enterococcus faecalis.
PN WO9850554-A2.
PD 12-NOV-1998.
XX 04-MAY-1998; 98WO-US08959.
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Bailey C, Choi GH, Bromocky J A, Kunsch CA;
PI WPI: 1999-070095/06.
XX P-PSDB: AAY00222.
DR WPI: 1999-070095/06.
XX New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection
XX Claim 1; Page 212; 301pp; English.

CC The present sequence represents a gene isolated from
CC Enterococcus faecalis. The present invention describes genes, proteins
CC and antigenic polypeptides isolated from E. faecalis. The proteins can
CC be used in vaccines for preventing or attenuating an infection caused
CC by a member of the Enterococcus genus in an animal. They can also be
CC used for detecting Enterococcus antibodies in a sample. The nucleotide

CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.

XX Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 888;

Best Local Similarity 100.0%; Pred. No. 1.4e+02; Mismatches 0; Indels 0; Gaps 0;

QY 2527 gctgtctgcaatgcag 2544
 |||||||
 DB 110 gctgtctgcaatgcag 127

RESULT 37

AAH67086 standard; DNA; 960 BP.

XX AAH67086;

XX 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 2121.

KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

RI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI; 2001-376931/40.

XX P-PSDB; AAC91867.

XX Novel polynucleotides derived from Corynebacterium, for identifying

XX mutation point of a gene, measuring expression of a gene, analysing

XX expression profile or pattern of a gene and identifying homologous gene

XX Claim 8; SEQ ID NO: 2121; 246pp + Sequence listing; English.

XX The present invention provides a number of nucleotide and protein

XX sequences from the Corynebacterium glutamicum Corynebacterium glutamicum. These

XX CC are useful for identifying the mutation point of a gene derived from a

XX CC mutant of Corynebacterium glutamicum, measuring expression amount and

XX CC analysing the expression profile or expression pattern of a gene derived

XX CC from Corynebacterium glutamicum, and identifying a homologue of a gene derived

XX CC from Corynebacterium glutamicum. Corynebacterium bacteria are useful for producing

XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

XX CC particularly L-lysine. The present sequence is a nucleic acid described

XX CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from the

XX CC European Patent Office.

XX CC Sequence 960 BP; 206 A; 235 C; 271 G; 248 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 960;
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.

XX Sequence 960 BP; 206 A; 235 C; 271 G; 248 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 960;

Best Local Similarity 100.0%; Pred. No. 1.4e+02; Mismatches 0; Indels 0; Gaps 0;

QY 3040 ttgttggtgcaattg 3057
 |||||||
 DB 394 ttgttggtgcaattg 411

RESULT 38

AAQ05870/C standard; cDNA; 992 BP.

XX AAQ05870;

XX 07-JAN-1991 (first entry)

DE Sequence encoding mammalian growth hormone receptor binding protein.

KM Ovine placental lactogen; ds;

XX Ovis ammon aries.

PN BP386979-A.

PD 12-SEP-1990.

PF 05-MAR-1990; 90EP-0302322.

PR 06-MAR-1989; 89US-0319585.

PR (GENT) GENENTECH INC.

XX Wood WI, Colosi PC;

PI WPI; 1990-276995/37.

DR P-PSDB; AAR06643.

XX DNA sequence encoding growth-hormone-receptor-binding protein -

XX useful for increase of mammalian meat and milk prodn. and

XX decrease in fat content.

XX Claim 14; Fig 3; 22pp; English.

XX Receptor binding protein may be used to increase meat and milk

XX CC production and decrease fat content especially in sheep and goats.

XX CC Sequence shows an identity of 26% with human growth hormone and 67%

XX CC to ovine placental lactogen.

XX CC Sequence was obtained using probe described in AAQ05869.

XX Sequence 992 BP; 308 A; 237 C; 195 G; 252 T; 0 other;

XX Query Match 0.6%; Score 18; DB 11; Length 992;

XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;

XX Mismatches 0; Indels 0; Gaps 0;

XX QY 535 ttgaagggaattgcga 552

XX DB 236 ttgaagggaattgcga 219

XX RESULT 39

XX AA252527/C

XX ID AA252527 standard; cDNA; 1034 BP.

XX AC AA252527;

XX XX

DT 29-FEB-2000 (first entry)
XX Human secreted protein clone ya66_1 nucleotide sequence SEQ ID NO:105.
DE
XX Human; secreted protein; immunostimulatory; haemostatic; cytokine;
KW proliferative; differentiative; chemotactic; chemokinetic; vaccine;
KM thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
XX gene therapy; ss.
OS Homo sapiens.
XX
PN WO958642-A2.
XX
PD 18-NOV-1999.
XX
PE 14-MAY-1999; 99WO-US10843.
XX
PR 14-MAY-1998; 98US-0085472.
PR 17-AUG-1998; 98US-0096824.
PR 11-SEP-1998; 98US-0099843.
PR 15-SEP-1998; 98US-0099950.
PR 29-SEP-1998; 98US-0100424.
PR 09-OCT-1998; 98US-0102329.
PR 11-DEC-1998; 98US-0103615.
PR 14-DEC-1998; 98US-0111799.
PR 31-DEC-1998; 98US-0112159.
PR 10-FEB-1999; 99US-0248059.
PR 06-APR-1999; 99US-0287150.
PR 13-MAY-1999; 99US-0311021.
XX
PA (GEMV) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ;
XX WPI: 2000-053095/04.
DR P-PSDB; AAY73442.
XX
PT Novel polynucleotides and proteins having biological activities which
PT make them suitable for treating, preventing or ameliorating medical
PT conditions in humans or animals -
XX
PS Claim 114; Page 653; 730pp; English.
XX
XX The present invention describes human secreted proteins encoded by
CC polynucleotides obtained from adult testes, foetal brain, adult brain,
CC brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC cDNA libraries. The polynucleotides and proteins are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals.
CC Suggested activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy. Therapeutic compositions
CC AA25281 encode human secreted proteins, and AAY7390 to AAY73500
CC represent human secreted proteins, given in the present invention.
XX
SQ Sequence 1034 BP; 305 A; 228 C; 222 G; 276 T; 3 other;

Query Match 0.6%; Score 18; DB 21; Length 1034;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 ctgtcgaagaagaccac 467
|
Db 168 CTGTGAGAGAGACCA 151

RESULT 40
AAAF71387
ID AAF71387 standard; DNA; 1083 BP.
XX
AC AAF71387;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:55.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100844-A2.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 27-AUG-1999; 99DE-1033005.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habernauer G;
XX WPI: 2001-061975/07.
DR P-PSDB; AAB79270.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 3; Page 217-219; 1246pp; English.
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar

CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB9243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).

Sequence 1083 BP; 241 A; 261 C; 308 G; 273 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 22; Length 1083;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3040 ttgttggtggtgcaattg 3057

DB 494 ttgttggtggtgcaattg 511

RESULT 41

AAZ42236
 ID AAZ42236 standard; cDNA; 1205 BP.

XX AAZ42236;

DT 31-JAN-2000 (first entry)

DE Human normal bladder tissue cDNA derived EST 115.

KW Human; bladder; treatment; EST; expressed sequence tag; cytostatic;

XX cancer; gene therapy; ss.

XX Homo sapiens.

PN DE19818620-A1.

XX 28-OCT-1999.

PF 21-APR-1998; 98DE-1018620.

PR 21-APR-1998; 98DE-1018620.

PA (META-) METAGEN CBS GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI: 1999-602416/52.

DR P-PSDB; AAY60559, AAY60560, AAY60561.

PT New polypeptides and their nucleic acids, useful for treatment of

XX bladder tumour and identification of therapeutic agents -

PS Claim 3; Page 235; 366pp; German.

CC This invention describes novel polypeptide fragment sequences (I) and
 CC their encoding nucleic acids (II) which are highly expressed in normal
 CC bladder tissue and have cytostatic activity. (II) are used for
 CC recombinant expression of (I) and to isolate complete genes. (I) are
 CC used to identify agents suitable for the treatment of bladder tumours,
 CC to directly treat this form of cancer (including expression from gene
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)
 CC is also used for the generation of specific antibodies. (II) are

CC identified by assembling ESTs (expressed sequence tags) from a
 CC particular tissue type before comparison of expression patterns. This
 CC allows a significantly longer fragment of the gene to be revealed, and
 CC therefore reduces the number of failures because of ESTs from different
 CC libraries representing different parts of the same unknown gene
 CC distorting the estimated frequency of occurrence in a particular tissue.
 CC AAZ42122-242248 represent EST fragments derived from a human normal
 CC bladder tissue cDNA library which encode the protein fragments
 CC represented in AAY60329-Y60591.

Sequence 1205 BP; 274 A; 349 C; 338 G; 244 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 20; Length 1205;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2016 actctgaggtcacactc 2033

DB 714 actctgaggtcacactc 731

RESULT 42

AAAN80316
 ID AAAN80316 standard; DNA; 1250 BP.

XX AAAN80316;

DT 10-JAN-1991 (first entry)

DE Transcription control region of gene encoding serine protease

DE (SP) of human myeloid cell origin.

KW Serine protease; myeloid cell; intravascular coagulation treatment; ss.

XX Homo sapiens.

PN WO8806621-A.

PD 07-SEP-1988.

PF 26-FEB-1988; 88WO-JP00205.

PR 09-SEP-1987; 87JP-0225540.

PR 05-MAR-1987; 87JP-0050676.

XX (TORA) TORAY IND INC (AOKI/).

PI Aoki Y, Okano K, Naruto M, Shimizu H, Nakamura H;

DR WPI: 1988-271161/38.

PT Serine protease of human myeloid cell origin -

PS useful in disseminated intravascular coagulation treatment

XX Disclosure; Fig 15; 52pp; Japanese.

CC SP DNA is selected and cloned from a human myeloid cell library,
 CC inserted in a suitable vector and expressed in a transformant of a
 CC suitable organism such as E. coli HMS-174 or HB-101 or a yeast. It is new
 CC so are a SP of human myeloid cell origin; precursors of SP having an
 CC N-terminal splittable or signal peptide; and a transcription-controlling
 CC DNA sequence required for expression of the gene. SP has antithrombotic
 CC activity, esp. useful in disseminated intravascular coagulation. By
 CC expression of the protease gene in a suitable transformant organism SP
 CC can be obtained in large quantity.

Sequence 1250 BP; 305 A; 340 C; 385 G; 220 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 9; Length 1250;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 tgggtgaagtgaattcc 1098
 |||||
Db 43 tgggtgaagtgaattcc 60

RESULT 43
AAC40012
ID AAC40012 standard; DNA: 1473 BP.
XX
AC AAC40012;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26727.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

XX	Ctenocephalides felis.
OS	
XX	WO9821324-A1.
PN	
PD	22-MAY-1998.
XX	
PF	10-NOV-1997; 97WO-US20598.
XX	
PR	12-NOV-1996; 96US-0747221.
PA	(HESK-) HESKA CORP.
XX	
PI	Brandt KS, Silver GM, Wisniewski N;
DR	WPI: 1998-297929/26.
DR	P-PSDB; AAM57852.
XX	
PT	New nucleic acid encoding carboxylesterase(s) from fleas - useful,
PT	e.g. in vaccines, for preventing infestation by hematophagous
PT	ectoparasites, particularly on cats and dogs
XX	
S)	Claim 1; Page 113-115; 23Opp; English.
XX	
CC	This sequence encodes the flea esterase protein, nFE51515 (the
CC	complementary strand is shown in AAV40738), of the invention. When
CC	administered to animals, the protein induces a protective immune
CC	(antibody) response against carboxylesterase (CE), so they, or
CC	compositions containing CE proteins, are used' therapeutically or as
CC	vaccines, to protect particularly mammals and birds, specifically cats
CC	and dogs, against hematophagous ectoparasite infestation (HEP)',
CC	specially fleas. More generally the compositions can be used to treat
CC	arthropods generally, including pests of agricultural crops, trees,
CC	stored goods etc., also those that are vectors of disease. Fragments of
CC	the DNA can be used as probes and primers for identification or
CC	production of nucleic acid. Antibodies against the protein can be used
CC	for passive immunisation: to screen expression libraries: to isolate the
CC	protein and to target cytotoxic compounds to HEP. The compounds
CC	containing CE are effective against both adult and larval stages; they
CC	target CEs, including juvenile hormone, that are involved in
CC	development, metamorphosis, feeding, digestion and reproduction.
SO	
Sequence	1515 BP; 504 A; 268 C; 314 G; 428 T; 1 other:
Query Match	0.6%; Score 18; DB 19; Length 1515;
Best Local Similarity	100.0%; Pred. No. 1,4e+02;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1002 ctccaattcgaatcc 1019
b	
363 CTCCAATTTCAGAAATCC	346
RESULT 46	
AAV40738	
ID	AAV40738 standard; cDNA: 1515 BP.
AC	AAV40738;
XX	
DJ	23-SEP-1998 (first entry)
XX	
DE	C. felis esterase, nFE51515, coding sequence complementary strand.
KW	Esterase; flaa; protective immune response; carboxylesterase; arthropod;
KM	hematophagous ectoparasite infestation; nFE51515; ds.
XX	
OS	Ctenocephalides felis.
XX	
PN	WO9821324-A1.
XX	
PD	22-MAY-1998.
XX	

PF 10-NOV-1997; 97WO-US20598.
 XX
 XX 12-NOV-1996; 96US-0747221.
 XX
 XX (HESK-) HESKA CORP.
 XX
 PI Brandt KS, Silver GM, Wisniewski N;
 XX
 DR WPI: 1998-297929/26.
 XX
 XX New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
 PT e.g. in vaccines, for preventing infestation by haematophagous
 PT ectoparasites, particularly on cats and dogs
 XX
 PS Claim 1; Page 115-116; 230pp; English.
 XX
 XX This sequence is the complementary strand of the DNA encoding the flea
 CC esterase protein, nFE51515 (see AAV40737 for coding strand), of the
 CC invention. When administered to animals, the protein induces a
 CC protective immune (antibody) response against carboxylesterase (CE), so
 CC they, or compositions containing CE proteins, are used, therapeutically
 CC or as vaccines, to protect particularly mammals and birds, specifically
 CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
 CC specifically fleas. More generally the compositions can be used to treat
 CC arthropods generally, including pests of agricultural crops, trees,
 CC stored goods etc., also those that are vectors of disease. Fragments of
 CC the DNA can be used as probes and primers for identification or
 CC production of nucleic acid. Antibodies against the protein can be used
 CC for passive immunisation; to screen expression libraries; to isolate the
 CC protein and to target cytotoxic compounds to HEP. The compounds
 CC containing CE are effective against both adult and larval stages; they
 CC target CE, including juvenile hormone, that are involved in
 CC development, metamorphosis, feeding, digestion and reproduction.
 XX
 SO Sequence 1515 BP; 428 A; 314 C; 268 G; 504 T; 1 other:

 Query Match 0.6%; Score 18; DB 19; Length 1515;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

 QY 1002 ctccaattcagaatcc 1019
 ||||||||||||||||
 Db 1153 ctccaattcagaatcc 1170

 RESULT 47
 AAF21032
 ID AAF21032 standard; DNA; 1561 BP.
 XX
 AC AAF21032;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 XX Human low adenosine antisense oligonucleotide related sequence #2599.
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antialstatic; analgesic; hypotensive; cyostatic;
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200062736-A2.
 XX
 XX 26-OCT-2000.
 XX

```

PE 24-MAR-2000; 2000OWO-US08020.
XX
XX 06-APR-1999; 99US-0127958.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX
XX (NYCE/) NYCE J W.
XX
XX
XX NYce JW;
XX
XX WPI; 2000-679539/66.
XX
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -
XX
XX
XX Disclosure: Page 840; 1592pp; English.
XX
XX
XX The present invention describes low adenosine (A) content antisense
XX oligonucleotides and compositions (I) comprising them. In the antisense
XX oligonucleotides the A is replaced by a 'universal' or alternative base.
XX (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
XX immunosuppressive, antisthmatic, hypotensive and cytostatic activities.
XX The antisense oligonucleotides and (I) can be used to down-regulate the
XX expression and or activity of target polypeptides associated with
XX lung/respiratory disorders and malignancies, such as stimulating and
XX activating peptide factors and transmitters, transcription factors,
XX immunoglobulins and antibodies, antibody receptors, cytokines and
XX chemokines, endogenously produced specific and non-specific enzymes,
XX binding proteins, adhesion molecules and their receptors, cytokine and
XX chemokine receptors, adenosine receptors, bradykinin receptors, central
XX nervous system (CNS) and peripheral nervous and non-nervous system
XX receptors, CNS and peripheral nervous and non-nervous system peptide
XX transmitters, defensins, growth factors, vasoactive peptides and
XX receptors, binding proteins and malignancy associated proteins. The
XX antisense oligonucleotides may be used in this way to treat disorders
XX including respiratory obstruction (especially pulmonary obstruction
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
XX and/or surfactant hypoproduction which are associated with a disease or
XX condition selected from pulmonary vasoconstriction, inflammation,
XX allergies, asthma, impeded respiration, respiratory distress syndrome
XX (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
XX hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
XX pulmonary transplantation rejection, pulmonary infections, bronchitis,
XX and/or cancer. AAF19434 to AAF21243 represent human polynucleotide
XX fragments and antisense oligonucleotides used in the exemplification of
XX the present invention.
XX
XX
XX Sequence 1561 BP; 354 A; 445 C; 485 G; 277 T; 0 other:
XX
XX
XX Query Match 0.6%; Score 18; DB 21; Length 1561;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1081 tgggtgaaagtgaagtc 1098
XX ||||||||||||||||
XX Db 142 tgggtgaaagtgaagtc 159
XX
XX
XX RESULT 48
XX AAA34910
XX ID AAA34910 standard; DNA; 1561 BP.
XX
XX AC AAA34910;
XX
XX XX 28-JUL-2000 (first entry)
XX
XX DE Human adenosine receptor related polynucleotide SEQ ID NO:2599.
XX
XX XX Human: adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
XX allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
XX antiallergic; antisthmatic; cytostatic; analgesic; impaired airway;

```

KM	Lung disease: ischaemic condition; pulmonary vasoconstriction; asthma;
KM	respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KM	pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KM	cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200009525-A2.
XX	
PD	24-FEB-2000.
XX	
PF	03-AUG-1999; 99WO-US17712.
XX	
PR	03-AUG-1998; 98US-0095212.
XX	
PA	(UYEC-) UNITV EAST CAROLINA.
XX	
PI	Nyce JW;
XX	
DR	WPI: 2000-205971/18.
XX	
PT	New antisense oligonucleotides useful for treating e.g. pulmonary
PT	vasoconstriction, inflammation, allergies, asthma, hypertension,
PT	bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT	cancers
PS	
XX	Disclosure; Page 769; 1343pp; English.
XX	
CC	The present invention describes a new composition comprising an
CC	antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC	targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC	inflammation. The ON can have antiinflammatory, antiallergic,
CC	antithrombotic, cytostatic and analgesic activities. The compositions are
CC	useful for the treatment of diseases associated with inflammation,
CC	impaired airways, including lung disease and diseases whose secondary
CC	effects afflict the lungs of a subject. They can be used for treating
CC	e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
CC	asthma, impaired respiration, respiratory distress syndrome, pain, cystic
CC	fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC	pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC	carcinomas, and cancers which may metastasize to the lungs, including
CC	breast and prostate cancer. The reduction of the adenosine content of
CC	the ONs reduces side effects. The A-containing ONs break down with the
CC	release of deoxyadenosine which activates adenosine receptors causing
CC	bronchoconstriction and inflammation. AAA32313 to AAA5312 represent the
CC	nucleotide sequences given in the sequence listing from the present
CC	invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC	185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC	differ from the previously named sequences. SEQ ID NO:11 to 1680
CC	(AAA32323 to AAA33992) are specifically claimed ONs from the present
CC	invention. N.B. Sequences given in the disclosure of the present
CC	invention do not match up with their corresponding SEQ ID NO: sequences
CC	given in the sequence listing.
XX	
SO	Sequence 1561 BP, 354 A; 445 C; 485 G; 277 T; 0 other;
XX	
Query Match	0.6%; Score 18; DB 21; Length 1561;
Best Local Similarity	100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1081 tgggtgaagtgagttcc 1098
DB	
142 tgggtgaagtgagttcc 159	
RESULT 49	
AAH13702/C	
ID	AAH13702 standard; cDNA; 1579 BP.
XX	
XX	AAH13702;
XX	
DT	26-JUN-2001 (first entry)

```

XX Human cDNA sequence SEQ ID NO:10578.
DE
XX
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
KW
XX
XX Homo sapiens.
OS
XX EPI074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PE
XX
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 10578; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 1579 BP; 468 A; 274 C; 283 G; 554 T; 0 other.
SQ

```

```

Query Match 0.6%; Score 18; DB 22; Length 1579;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 36 atgaattcataaagac 53
    ||||||||||||||||
Db 1229 ATGATTCATCAAGAC 1212

```

```

RESULT 50
AAQ12528/C
ID AAQ12528 standard; DNA; 1611 BP.
XX
XX AC . AAQ12528;

```

```

XX 25-SEP-1991 (first entry)
DT
XX
XX Thymidylate phosphorylase.
DE
XX
XX Deoxyribonucleic acid; thymidine; dTMP; pyrimidine; ss.
KW
XX
XX Bacillus subtilis CMG356 (ATCC 33234).
OS
XX
XX Key Location/Qualifiers
FT CDS 136..857
FT /tag=a
FT /product=dTMPase
XX
XX WO9109130-A.
PN
XX
XX 27-JUN-1991.
PD
XX
XX 05-DEC-1990; 90WO-US06993.
PF
XX
XX 08-DEC-1989; 89US-0448158.
PR
XX
XX (CHEM-) CHEMGEN CORP.
PA
XX
XX Mc Danliss RL, Anderson DM;
XX
XX WPI: 1991-208156/28.
DR
XX P-PSDB; AAR12555.
DR
XX
XX Microorganism contg. deoxyribonucleic acid - encoding enzyme
PT causing accumulation of pyrimidine deoxyribonucleoside in
PT recoverable amts.
PT
XX
XX Disclosure: Fig 7(a-c); 79pp; English.
XX
XX The sequence, comprising flanking DNA sequences obtd. from a PSI
PS bacteriophage, encodes an enzyme capable of converting a thymidine
PS deoxyribonucleoside monophosphate to a thymidine deoxyribonucleoside.
CC
CC The sequence may be used together with metabolic mutations of
CC heterologous DNA, encoding metabolic enzymes, to engineer cultured
CC cells to express thymidine deoxyribonucleosides (TdNs) in recoverable
CC amts., providing a fermentation source of TdNs.
CC See also AAQ12526-27.
XX
XX Sequence 1611 BP; 683 A; 171 C; 244 G; 513 T; 0 other;
SQ

```

```

Query Match 0.6%; Score 18; DB 12; Length 1611;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 343 aggaattaaagactgt 360
    ||||||||||||||||
Db 979 AGGATTTAAGCAGCTGT 962

```

```

RESULT 51
AAA47151/C
ID AAA47151 standard; DNA; 1611 BP.
XX
XX AAA47151;
AC
XX
XX 03-OCT-2000 (first entry)
DT
XX
XX DNA encoding a serine protease inhibitor protein.
DE
XX
XX Serine protease inhibitor; green-lipped mussel; anti-thrombin;
KW divalent metal cation binding activity; dietary supplement;
KW anticoagulant; ss.
XX
XX Perna canaliculus.
OS
XX
XX Key Location/Qualifiers
FH

```

```

AAQ12528/C
ID AAQ12528 standard; DNA; 1611 BP.
XX
XX AC . AAQ12528;

```



```
FT CDS 1..1494
FT /*tag= a
FT /product= "serine protease inhibitor"
FT polyA_site 1557..1563
FT /*tag= b
PN WO200039165-A1.
XX 06-JUL-2000.
XX
XX 23-DEC-1999; 99WO-NZ00227.
XX
XX 23-DEC-1998; 98NZ-0333568.
XX
XX 23-JUL-1999; 99NZ-0336906.
XX
XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX
XX Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
XX
XX WPI: 2000-452375/39.
XX
XX P-PSDB: AAY93750.
XX
XX New Perna canaliculus serine protease inhibitor protein exhibiting
XX anti-thrombin activity and divalent metal cation binding activity,
XX useful as an anticoagulant agent and as a dietary supplement -
XX
XX Claim 11; Page 10-11; 44pp; English.
XX
XX The present sequence encodes a serine protease inhibitor
XX protein. The protein is isolated from the green-lipped mussel
XX (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
XX and divalent metal cation binding activity. The serine protease
XX inhibitor protein has a molecular weight of about 55 kilo Daltons.
XX The protein, and its fragments, are useful in medicaments, in food,
XX as dietary supplements or as bioremediation agents. In the dietary
XX supplements, the protein is associated with or bound to at least one
XX divalent cation (such as calcium, magnesium or zinc) of dietary
XX significance. The proteins or their fragments are also useful as
XX anticoagulant agents.
XX
XX Sequence 1611 BP; 499 A; 348 C; 360 G; 402 T; 2 other;
XX
XX
XX Query Match 0.6%; Score 18; DB 21; Length 1611;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1051 tcttggtagtcacactt 1068
XX ||||||||||||||||
XX DB 767 TCTTTGTGTCATCACTT 750
XX
XX RESULT 52
XX AAA15908/C
XX ID AAA15908 standard; CDNA: 1619 BP.
XX
XX AAA15908;
XX
XX 12-JUN-2000 (first entry)
XX
XX Human protein clone HP10195 full length coding sequence.
XX
XX Human protein: hydrophobic domain; nutritional source: haematopoiesis;
XX cytokine production; cell proliferation; cell differentiation;
XX immune deficiency; infectious disease; autoimmune disorder; asthma;
XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
XX allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
XX nervous system disorder; Alzheimer's disease; Parkinson's disease;
XX Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
XX systemic cytokine damage; tissue differentiation; contraceptive; stroke;
XX coagulation disorder; myocardial infarction; inflammatory condition;
XX septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
XX nephritis; therapy; ss.
```

```
XX Homo sapiens.
OS
XX
XX WO200005367-A2.
XX
XX 03-FEB-2000.
XX
XX 22-JUL-1999; 99WO-JP03929.
XX
XX 24-JUL-1998; 98JP-0208820.
XX
XX 07-AUG-1998; 98JP-0224105.
XX
XX 25-AUG-1998; 98JP-0238116.
XX
XX 09-SEP-1998; 98JP-0254736.
XX
XX 29-SEP-1998; 98JP-0275505.
XX
XX (SAGA ) SAGAMI CHEM RES CENT.
XX (PROT-) PROTEGENE INC.
XX
XX Kato S, Kimura T;
XX
XX WPI: 2000-182694/16.
XX
XX P-PSDB: AAY94850.
XX
XX Novel human proteins having hydrophobic domains useful for treating
XX osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
XX multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
XX Claim 4; Page 192-194; 351pp; English.
XX
XX This sequence encodes a human protein of the invention, which has
XX hydrophobic domains. The DNA sequences can be used as a probe or as a
XX genetic marker. The protein can also be used as a marker, and to identify
XX potential genetic disorders. The DNA and protein can also be used as
XX nutritional sources or supplements. The protein exhibits cytokine, cell
XX proliferation, cell differentiation activities and induces production of
XX other cytokines in certain cell populations. The protein also exhibits
XX immune stimulating or immune suppressing activity. It can be used in the
XX treatment of various immune deficiencies and disorders, and to treat
XX infectious diseases caused by viral, bacterial, fungal or other
XX infections. The protein is also used for treating autoimmune disorders
XX such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
XX arthritis. It is also useful in the treatment of allergic reactions and
XX conditions such as asthma, and in immune suppression after organ
XX transplantation. The protein is useful in regulation of haematopoiesis
XX and consequently in the treatment of myeloid or lymphoid cell
XX deficiencies. It is also used in compositions for tissue growth or
XX regeneration. The protein is also used in the treatment of osteoporosis
XX or osteoarthritis and in the treatment of periodontal disease and other
XX tooth repair processes. The protein is used in the treatment of nervous
XX system disorders such as Alzheimer's disease, Parkinson's disease, and
XX Huntington's disease. They are useful for protection or regeneration and
XX treatment of lung or liver fibrosis, reperfusion injury in various
XX tissues, and conditions resulting from systemic cytokine damage. They are
XX also used for promoting or inhibiting tissue differentiation. They are
XX also used as contraceptives since they exhibit activin or inhibin related
XX activities and as a fertility inducing therapeutic. They are used for
XX treating various coagulation disorders and in treatment and prevention of
XX conditions resulting from coagulation activities e.g. myocardial
XX infarction or stroke. They also acts as receptors, receptor ligands or
XX inhibitors or agonists of receptor/ligand interactions. They are used to
XX treat inflammatory conditions such as septic shock, sepsis, ischemia
XX reperfusion injury, arthritis, and nephritis. They can be used to
XX prevent tumours.
XX
XX Sequence 1619 BP; 444 A; 359 C; 363 G; 453 T; 0 other;
XX
XX
XX Query Match 0.6%; Score 18; DB 21; Length 1619;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 36 atgaattcataagac 53
XX ||||||||||||||||
```

Db 1277 ATGATTTCATTAAGAC 1260

RESULT 53
AAV40760/c
ID AAV40760 standard; cDNA; 1650 BP.

XX
XX AAV40760;
XX
XX 23-SEP-1998 (first entry)
XX
XX C. felis esterase, nFE51650, coding sequence.
DE
XX
XX Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE51650; ds.
XX
XX Ctenocephalides felis.
OS
XX
XX MO9821324-A1.
PN
XX
XX 22-MAY-1998.
PD
XX
XX 10-NOV-1997; 97WO-US20598.
PR
XX 12-NOV-1996; 96US-0747221.
PA (HESK-) HESKA CORP.
PI Brandt KS, Silver GM, Wisniewski N;
XX
XX WPI: 1998-297929/26.
DR P-PSDB; AAM57865.
XX
XX
XX New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PI ectoparasites, particularly on cats and dogs
XX
XX
XX Claim 1; Page 188-191; 230pp; English.

XX
XX This sequence encodes the flea esterase protein, nFE51650 (the
CC complementary strand is shown in AAV40761), of the invention. When
CC administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.

XX
XX Sequence 1650 BP; 542 A; 284 C; 341 G; 482 T; 1 other;

XX
XX Query Match 0.6%; Score 18; DB 19; Length 1650;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
|||||
Db 498 CTCCAATTCGAAATCC 481

RESULT 54
AAV40761
ID AAV40761 standard; cDNA; 1650 BP.
XX

AC AAV40761;
XX
XX 23-SEP-1998 (first entry)
DT
XX
XX C. felis esterase, nFE51650, coding sequence complementary strand.
DE
XX
XX Esterase; flea; protective immune response; carboxylesterase; arthropod;
KW haematophagous ectoparasite infestation; nFE51650; ds.
XX
XX Ctenocephalides felis.
OS
XX
XX MO9821324-A1.
PN
XX
XX 22-MAY-1998.
PD
XX
XX 10-NOV-1997; 97WO-US20598.
PF
XX 12-NOV-1996; 96US-0747221.
PR
XX
XX (HESK-) HESKA CORP.
PA
XX
XX Brandt KS, Silver GM, Wisniewski N;
PI
XX
XX WPI: 1998-297929/26.
DR
XX
XX
XX New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PI ectoparasites, particularly on cats and dogs
XX
XX
XX Claim 1; Page 191-192; 230pp; English.

XX
XX This sequence is the complementary strand of the DNA encoding the flea
CC esterase protein, nFE51650 (see AAV40760 for coding strand), of the
CC invention. When administered to animals, the protein induces a
CC protective immune (antibody) response against carboxylesterase (CE), so
CC they, or compositions containing CE proteins, are used, therapeutically
CC or as vaccines, to protect particularly mammals and birds, specifically
CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.

XX
XX Sequence 1650 BP; 482 A; 341 C; 284 G; 542 T; 1 other;

XX
XX Query Match 0.6%; Score 18; DB 19; Length 1650;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ctccaattcagaatcc 1019
|||||
Db 1153 ctccaattcagaatcc 1170

RESULT 55
AAV71287/c
ID AAV71287 standard; DNA; 1723 BP.
XX
XX AAV71287;
AC
XX
XX 12-AUG-1999 (first entry)
DT
XX
XX Human vesicular binding protein nucleotide sequence.
DE
XX
XX VSBP; fibroblast; Aplysia; VAMP; vesicle associated membrane protein;
KW synaptobrevin binding protein; cancer; inflammation; ss.

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 568..1278
FT /tag= a
FT /product= "MYBP"
XX
XX PN MO9851797-A1.
XX
XX PD 19-NOV-1998.
XX
XX PF 15-MAY-1998; 98WO-US10225.
XX
XX PR 15-MAY-1997; 97US-0857213.
XX
XX PA (INCY-) INCYTE PHARM INC.
XX
XX I Bandman O, Hawkins PR, Murry LE;
XX
XX DR WPI: 1999-045229/04.
XX
XX P-PSDB; AAW86813.
XX
XX PT Human vesicle binding protein - useful for the diagnosis, treatment
XX and prevention of cell proliferation disorders and inflammation
XX
XX PS Claim 5; Fig 1A-D; 58pp; English.
XX
XX CC The present sequence represents a nucleic acid encoding a purified human
XX vesicle binding protein (designated MYBP). This was obtained by standard
XX cloning and screening procedure, where it was first isolated in Incyte
XX clone 148415 from a normal fibroblast library (FIBRNGT01). A comparison
XX of the MYBP with the Aplysia vesicle-associated membrane protein
XX (VAMP)/synaptobrevin binding protein showed that they shared 44%
XX identity. MYBP and its products can be used to diagnose, treat or prevent
XX disorders of cell proliferation (e.g. cancer) or inflammation. Sequences
XX complementary to MYBP can be used to detect MYBP in samples. Abs are used
XX to diagnose conditions associated with MYBP or in assays to monitor
XX patients being treated with MYBP or agonists/antagonists.
XX
XX SQ Sequence 1723 BP; 548 A; 356 C; 325 G; 468 T; 26 other;

Query Match 0.6%; Score 18; DB 20; Length 1723;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

36 atgaattcataaagac 53
|||||
Db 1540 ATGAATTCATTAAGAC 1523

RESULT 56
AA067601
ID AA067601 standard; cDNA; 1800 BP.
XX
XX AC AA067601;
XX
XX DT 20-NOV-1994 (first entry)
XX
XX DE Retinoblastoma protein Ap4.
XX
XX KM Retinoblastoma-associated polypeptide; RAP; Ap12; Ap4;
XX transcription factor; EF2; cell cycle; ss.
XX
XX OS Homo sapiens.
XX
XX PN MO9412521-A.
XX
XX PD 09-JUN-1994.
XX
XX PF 19-NOV-1993; 93WO-US11310.
XX

```

```

PR 20-NOV-1992; 92US-0979156.
XX
XX PA (TEXA ) UNIV TEXAS SYSTEM.
XX
XX PI Lee W, Shan B;
XX
XX DR WPI: 1994-200185/24.
XX
XX PT Nucleic acid encoding retinoblastoma-associated polypeptide(s) -
XX used for producing prods. for use in studying cell cycling and
XX PT dysregulated cell growth.
XX
XX PS Disclosure; Page 37-38; 77pp; English.
XX
XX CC cDNA encoding a retinoblastoma (RB) associated protein (RAP), Ap12,
XX was cloned by direct screening of cDNA expression libraries using
XX purified RB protein as probe. Ap12 showed transcription factor E2F
XX and RB binding activities. The 5' and 3' cDNA sequences were
XX determined for Ap2 (AA067395, AA067596), Ap8 (AA067597, AA067598) and
XX Ap15 (AA067599, AA067600), and full-length encoding sequences for Ap4
XX (AA067601) and Ap10 (AA067601), which also encoded RB-associated
XX proteins.
XX
XX SQ Sequence 1800 BP; 747 A; 333 C; 372 G; 348 T; 0 other;

Query Match 0.6%; Score 18; DB 15; Length 1800;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2831 aaagaacctctgaaaa 2848
|||||
Db 307 aaagaacctctgaaaa 324

RESULT 57
AAF25374
ID AAF25374 standard; cDNA; 1878 BP.
XX
XX AC AAF25374;
XX
XX DT 15-MAY-2001 (first entry)
XX
XX DE Nucleotide sequence of the Arabidopsis SGS3 polypeptide.
XX
XX KM SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX viral resistance; resistance; fatty acid content; protein content; ss.
XX
XX OS Arabidopsis thaliana.
XX
XX FH Key Location/Qualifiers
XX CDS 1..1878
XX /tag= a
XX /product= "SGS3"

WO200105951-A2.
XX
XX PD 25-JAN-2001.
XX
XX PF 13-JUL-2000; 2000WO-FR02052.
XX
XX PR 16-JUL-1999; 99FR-0009417.
XX 26-JAN-2000; 2000FR-0001006.
XX
XX PA (AVET ) AVENTIS CROPS SCIENCE SA.
XX (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX PI Beclin C, Elmayer T, Vaucheret H;
XX
XX DR WPI: 2001-159528/16.
XX
XX P-PSDB; AAB31796.
XX
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus

```

PT resistance in plants and, when inhibited, for increasing transgene
expression -
XX
XX
PS Claim 1; Page 32-35; 36pp; French.
XX
CC The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.
CC CC (The SGS3 gene is essential for post-transcriptional inactivation
CC of degradation of RNA) and for resistance to viruses. Overexpression
CC of SGS3 results in plants with increased resistance to viruses,
CC while inactivation of SGS3 in transgenic plants (e.g. by expressing
CC antisense RNA, by mutation or by homologous recombination) increases
CC the level of the transgene product. This product may e.g. impart
CC resistance (to herbicide, insects or pathogens), alter contents
CC of essential fatty acids or proteins, or is pharmaceutically active,
CC e.g. an immunoglobulin or interferon.
XX
SQ Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 other;

Query Match 0.6%; Score 18; DB 22; Length 1878;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1020 aggtcgttgaggaaatc 1037
|||||
Db 1388 aggtcgttgaggaaatc 1405

RESULT 58
AAV82522
ID AAV82522 standard; DNA; 1926 BP.
XX
AC AAV82522;
XX
DT 17-MAR-1999 (first entry)
XX
DE Sphingomonas capsulata mature aminopeptidase I encoding DNA.
XX
KW Aminopeptidase; protein hydrolysate; glycine releasing; protease;
KW proteolaceous material; flavour; food; baking; animal feed additive;
KW palatability; hydrolysis; solubility; emulsifying; foaming; aroma;
KW mouthfeel; crust colour; ss.
XX
OS Sphingomonas capsulata.
XX
FH Key Location/Qualifiers
FT CDS 1..1491
FT /tag= a
XX
XX W09851163-A2.
XX
XX 19-NOV-1998.
XX
PE 15-MAY-1998; 98WO-US09998.
XX
XX 16-DEC-1997; 97US-0069719.
XX 16-MAY-1997; 97US-0857886.
XX 20-OCT-1997; 97US-0062893.
XX 16-DEC-1997; 97DK-0001465.
XX
XX (ASAH) ASAH CHEM IND CO LTD.
XX PA (NOVO) NOVO NORDISK BIOTECH INC.
XX PA (NOVO) NOVO-NORDISK AS.
XX
XX Blinkovsky A, Brown K, Byun T, Fujii M, Golightly E;
XX PI Kotod IV, Marumotac, Mathiansen TE;
XX
XX WPI: 1999-045177/04.
XX P-PSDB: AAV89587.
XX
PT Production of protein hydrolysate - using protease and enzyme that
PT releases glycine, useful as flavour improvers in foods and animal
PT feed additives

XX
PS Claim 7; Page 61; 84pp; English.
XX
CC A method has been developed for the production of protein hydrolysates
CC (PH) comprising reacting a protein with: (i) at least one polypeptide
CC with Gly-releasing activity; and (ii) at least one other protease so
CC that the amount of Gly produced is greater than when the protease is
CC used alone. PH are used to improve flavour of foods (e.g. baked goods)
CC and as animal feed additives. PH are preferably also enriched in Glu
CC (free and/or peptide bound), so have improved flavour and palatability.
CC Addition of a polypeptide with Gly-releasing activity increases the
CC degree of hydrolysis (or reduces the amount of enzyme needed) and
CC hydrolysates have better solubility, and emulsifying and foaming
CC properties. Baked goods containing them have improved aroma, mouthfeel
CC and crust colour. The present sequence encodes Sphingomonas capsulata
CC aminopeptidase I, which is used in the method of the invention.
XX
SQ Sequence 1926 BP; 362 A; 633 C; 628 G; 303 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 1926;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2127 gctgtgtgtgtgtgaagc 2144
|||||
Db 1242 gctgtgtgtgtgtgaagc 1259

RESULT 59
AAZ77504
ID AAZ77504 standard; CDNA; 1962 BP.
XX
AC AAZ77504;
XX
DT 10-APR-2000 (first entry)
XX
DE Human ovarian tumor CDNA library derived EST fragment 55.
XX
XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW gene therapy; treatment; ss.
XX
XX Homo sapiens.
XX
XX DE19817557-A1.
XX
XX 21-OCT-1999.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX PI WPI: 1999-591920/51.
XX DR P-PSDB: AAV76633, AAV76634, AAV76635.
XX
XX New nucleic acid sequences expressed in ovarian, and some other, cancer
XX tissues, and derived polypeptides, for treatment of ovarian cancer and
XX identification of therapeutic agents -
XX
XX Claim 3; Page 182-183; 310pp; German.

This invention describes novel nucleic acid (cDNA) sequences (A) which
have anticancer activity and are highly expressed in ovarian tumor
tissue (and some also in testis and breast cancer tissue). The products
of the invention can be used for gene therapy. (A) are used (i) for
CC recombinant expression of polypeptides (B) and (ii) to isolate complete
CC genes. (B) are used (i) to identify agents suitable for treatment of
CC ovarian cancer; (ii) directly for treating this form of cancer
CC (including expression from gene therapy vectors) and (iii) for generation

CC of specific antibodies. (A) are identified by assembling ESTs (expressed
CC sequence tags) from a particular tissue type before comparison of
CC expression patterns. This allows a significantly longer fragment of the
CC gene to be revealed, so should reduce the number of failures associated
CC with the fact that ESTs from different libraries may represent different
CC parts of the same unknown gene, distorting the estimated frequency of
CC occurrence in a particular tissue. AA277450-277572 represent the human
CC ovarian tumor cDNA library derived EST fragments described in the method
CC of the invention and encode the protein fragments represented in
CC AAV76505-Y76638.

XX
SQ Sequence 1962 BP; 596 A; 357 C; 389 G; 620 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 18; DB 20; Length 1962;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 atgaattcataaagac 53
|||||
Db 368 atgaattcataaagac 385

RESULT 60
AAV40735/C
ID AAV40735 standard; cDNA; 1982 BP.
XX
AC AAV40735;
XX
DT 23-SEP-1998 (first entry)
XX
DE C. felis esterase, nFE51982, coding sequence.
XX
KM Esterase; flea; protective immune response; carboxylesterase; arthropod;
KM haematophagous ectoparasite infestation; nFE51982; ds.
XX
OS Ctenocephalides felis.
XX
FH Key Location/Qualifiers
FT CDS 31..1520
FT /*tag= a
XX
PN WO9821324-A1.
XX
PD 22-MAY-1998.
XX
PF 10-NOV-1997; 97WO-US20598.
XX
PR 12-NOV-1996; 96US-0747221.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Silver GM, Wisniewski N;
XX
DR WPI; 1998-297929/26.
DR P-SDB; AAW57851.
XX
PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
XX
PS Claim 1; Page 106-109; 230pp; English.

CC This sequence encodes the flea esterase protein, nFE51982 (the
CC complementary strand is shown in AAV40736), of the invention. When
CC administered to animals, the protein induces a protective immune
CC (antibody) response against carboxylesterase (CE), so they, or
CC compositions containing CE proteins, are used, therapeutically or as
CC vaccines, to protect particularly mammals and birds, specifically cats
CC and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of

CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.

XX
SQ Sequence 1982 BP; 691 A; 304 C; 357 G; 629 T; 1 other;

Query Match
Best Local Similarity 100.0%; Score 18; DB 19; Length 1982;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1002 ctccaattcagaatcc 1019
|||||
Db 365 CTCCAATTCAGAAATCC 348

RESULT 61
AAV40736
ID AAV40736 standard; cDNA; 1982 BP.
XX
AC AAV40736;
XX
DT 23-SEP-1998 (first entry)
XX
DE C. felis esterase, nFE51982, coding sequence complementary strand.
XX
KM Esterase; flea; protective immune response; carboxylesterase; arthropod;
KM haematophagous ectoparasite infestation; nFE51982; ds.
XX
OS Ctenocephalides felis.
XX
PN WO9821324-A1.
XX
PD 22-MAY-1998.
XX
PF 10-NOV-1997; 97WO-US20598.
XX
PR 12-NOV-1996; 96US-0747221.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Silver GM, Wisniewski N;
XX
DR WPI; 1998-297929/26.
XX
PT New nucleic acid encoding carboxyl:esterase(s) from fleas - useful,
PT e.g. in vaccines, for preventing infestation by haematophagous
PT ectoparasites, particularly on cats and dogs
XX
PS Claim 1; Page 111-112; 230pp; English.

CC This sequence is the complementary strand of the DNA encoding the flea
CC esterase protein, nFE51982 (see AAV40735 for coding strand), of the
CC invention. When administered to animals, the protein induces a
CC protective immune (antibody) response against carboxylesterase (CE), so
CC they, or compositions containing CE proteins, are used, therapeutically
CC or as vaccines, to protect particularly mammals and birds, specifically
CC cats and dogs, against haematophagous ectoparasite infestation (HEP),
CC specifically fleas. More generally the compositions can be used to treat
CC arthropods generally, including pests of agricultural crops, trees,
CC stored goods etc., also those that are vectors of disease. Fragments of
CC the DNA can be used as probes and primers for identification or
CC production of nucleic acid. Antibodies against the protein can be used
CC for passive immunisation; to screen expression libraries; to isolate the
CC protein and to target cytotoxic compounds to HEP. The compounds
CC containing CE are effective against both adult and larval stages; they
CC target CEs, including juvenile hormone, that are involved in
CC development, metamorphosis, feeding, digestion and reproduction.

RESULT 64
AAZ96326/c
ID AAZ96326 standard; DNA: 2299 BP.
XX
AC AAZ96326;
XX
DT 10-APR-2000 (first entry)
XX
DE S. pneumoniae derived DNA from ORF #154.
XX
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
KM bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO9806734-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14436.
XX
PR 16-AUG-1996; 96US-0024022.
XX
PA (SMIR) SMITHKLINE BEECHAM CORP.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
DR WPI: 1998-159452/14.
DR P-PSDB; AAY85980, AAY85981.
XX
PT Streptococcus pneumoniae proteins and related DNA - useful for
PT screening compounds for antibacterial activity
XX
PS Claim 4; Page 188-189; 640pp; English.
XX
PS This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see
CC AAY87297-186182). The DNA, vectors and host cells described in the
CC method of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC particularly useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.
XX
SQ Sequence 2299 BP; 585 A; 514 C; 492 G; 708 T; 0 other;

Query Match 0.6%; Score 18; DB 19; Length 2299;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2765 caagcttggtgaataa 2782
|||||
DB 429 CAAGCTTGCGTGAATAA 412

RESULT 65
AAV42995/c
ID AAV42995 standard; DNA: 2299 BP.
XX
AC AAV42995;
XX
DT 09-NOV-1998 (first entry)
XX
DE Streptococcus pneumoniae polypeptide coding region.

XX
KM Polypeptide; ORF; open reading frame; infection; bacterial;
KM streptococcal; bacteremia; diagnosis; prophylaxis; ds.
XX
XX Streptococcus pneumoniae.
OS
FH Key Location/Qualifiers
FT CDS complement (388..1008)
FT /*tag= a
FT /note= "polypeptide"
XX
PN WO9823631-A1.
XX
PD 04-JUN-1998.
XX
PF 24-NOV-1997; 97WO-US21976.
XX
PR 27-NOV-1996; 96US-0031879.
XX
PA (SMIR) SMITHKLINE BEECHAM CORP.
PA (SMIR) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Reid RH, Zarfos PN;
XX
DR WPI: 1998-322654/28.
DR P-PSDB; AAW62719.
XX
PS Claim 1; Page 106; 181pp; English.
XX
PS The sequence is that of a Streptococcal polypeptide coding region.
CC The polypeptide can potentially be used for the diagnosis and
CC prevention of bacterial infections, especially SP infection.
CC It may be used for the treatment of diseases such as otitis media,
CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC empyema, endocarditis or infection of the cerebrospinal fluid.
XX
SQ Sequence 2299 BP; 585 A; 514 C; 491 G; 708 T; 1 other;

Query Match 0.6%; Score 18; DB 19; Length 2299;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2765 caagcttggtgaataa 2782
|||||
DB 429 CAAGCTTGCGTGAATAA 412

RESULT 66
AAW84399
ID AAW84399 standard; DNA: 3000 BP.
XX
AC AAW84399;
XX
DT 09-SEP-1999 (first entry)
XX
DE S. capsulata IF012533 aminopeptidase coding sequence.
XX
KW Aminopeptidase; amino acid removal; protein hydrolysatate production;
KW hydrolysis; flavour development; enzyme deactivation;
KW peptide sequence cleavage; post-translational processing;
KW precursor protein activation; ss.
XX
OS Sphingomonas capsulata.
XX
PN WO931226-A1.
XX
PD 24-JUN-1999.

```

XX PF 13-NOV-1998; 98WO-DK00495.
XX PR 15-MAY-1998; 98DK-0000670.
PR 16-DEC-1997; 97DK-0001465.
PR 16-DEC-1997; 97US-0069719.
XX PA (ASAH ) ASAH1 CHEM IND CO LTD.
PA (NOVO ) NOVO NORDISK BIOTECH INC.
XX PA (NOVO ) NOVO-NORDISK AS.
XX PI Blinkovsky A, Brown K, Byun TS, Fujii M, Klotz AV;
PI Matumoto C, Sioma A, Tang M;
XX DR WPI: 1999-418769/35.
DR P-PSDB: AAY22184.
XX PT New isolated aminopeptidase polypeptides used in, e.g. food industry
PT Claim 14; Fig 9; 84pp; English.
XX PS
XX CC This sequence encodes the Sphingomonas capsulata IF012533 aminopeptidase
CC of the invention. The aminopeptidase polypeptides catalyse the removal of
CC amino acids from the N-terminal end of peptides, oligopeptides or
CC proteins. They can be used in the production of protein hydrolysates for
CC enhancing the degree of hydrolysis and flavour development, particularly
CC in foods. They can also be used to deactivate enzymes. They can also be
CC used for specific cleavage of peptide sequences, e.g. to provide the
CC necessary post-translational processing to activate precursor proteins.
XX SQ Sequence 3000 BP; 528 A; 978 C; 978 G; 516 T; 0 other;

Query Match 0.6%; Score 18; DB 20; Length 3000;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2127 gctggtgtgctggaagc 2144
DB 1911 gctggtgtgctggaagc 1928

RESULT 67
AAH51776/C
ID AAH51776 standard; DNA; 3001 BP.
XX AC AAH51776;
XX DT 29-AUG-2001 (first entry)
XX CC Chromosome 13q31-q33 biallelic marker containing amplicon SEQ ID 188.
XX KW sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
XX KM biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX OS Homo sapiens.
XX PN WO200058510-A2.
XX PD 05-OCT-2000.
XX PF 30-MAR-2000; 2000WO-IB00435.
XX PR 30-MAR-1999; 99US-0126903.
PR 30-APR-1999; 99US-0131971.
PR 30-APR-1999; 99US-0132065.
PR 14-JUL-1999; 99US-0143928.
PR 27-JUL-1999; 99US-0145915.
PR 29-JUL-1999; 99US-0146452.
PR 29-JUL-1999; 99US-0146453.
PR 28-OCT-1999; 99US-0162288.
XX PA (GEST ) GENSET.

```

```

XX PI Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;
PI Essiloux L;
XX DR WPI: 2000-619082/59.
XX PT Polynucleotides comprising sequences from sbg1 and g35018 biallelic
XX markers are used for genotyping and detecting schizophrenia or bipolar
XX disorder and predisposition to these disorders -
XX PS Claim 2; Page 680-681; 737pp; English.
XX CC AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
CC human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
CC AAH62907 - AAH62915 represent CDNA human sbg1 CDNA sequences and protein
CC products. AAH51627 - AAH51631 and AAH62916 - AAH62918 represent g35018
CC CDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are
CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC amplicons which comprise biallelic markers located on the chromosome
CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers
CC are represented in the sequences by degenerate/undefined base codes. PCR
CC primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC the invention. The biallelic marker containing nucleotide sequences are
CC used to determine the identity of the nucleotide at a biallelic marker in
CC a sample DNA sequence. The nucleotide sequences may be labelled and used
CC for genotyping by determining the identity of a nucleotide at a region
CC D-related biallelic marker in a biological sample from single or multiple
CC subjects. By determining the frequency of a biallelic marker in a
CC population an association between a genotype and a trait, a haplotype and
CC a trait and a phenotype and a trait can be detected. The sequences can be
CC used to determine a predisposition to or early onset of schizophrenia or
CC bipolar disorder or a beneficial response to or side effects related to
CC treatment against schizophrenia or bipolar disorder.
XX SQ Sequence 3001 BP; 993 A; 463 C; 533 G; 1011 T; 1 other;

Query Match 0.6%; Score 18; DB 21; Length 3001;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1663 aagaaattctgaaagcca 1680
DB 181 AAGAAATTCCTGAAGCCA 164

RESULT 68
AAH51787/C
ID AAH51787 standard; DNA; 3001 BP.
XX AC AAH51787;
XX DT 29-AUG-2001 (first entry)
XX CC Chromosome 13q31-q33 biallelic marker containing amplicon SEQ ID 199.
XX KW sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
XX KM biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX OS Homo sapiens.
XX PN WO200058510-A2.
XX PD 05-OCT-2000.
XX PF 30-MAR-2000; 2000WO-IB00435.
XX PR 30-MAR-1999; 99US-0126903.
PR 30-APR-1999; 99US-0131971.
PR 30-APR-1999; 99US-0132065.

```


KW Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
KM detection; colon cancer cell line Km12L4-A; ss.
XX Homo sapiens.
OS
XX WO200018916-A2.
PN
XX 06-APR-2000.
PD
XX 23-SEP-1999; 99WO-US22226.
PE
XX 28-SEP-1998; 98US-0102161.
PR 28-SEP-1998; 98US-0102180.
PR 29-SEP-1998; 98US-0102380.
PR 08-OCT-1998; 98US-0103815.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HSE-) HSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Cirvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX WPI: 2000-293155/25.
DR
XX Polynucleotide library comprising 1079 defined sequences, useful in
PT the form of an array to detect cancer or susceptibility to cancer -
PT
XX Claim 1; Page 501-502; 502pp; English.
XX
CC The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (I);
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Km12L4-A CDNA library.
XX
XX Sequence 3529 BP; 1248 A; 778 C; 692 G; 803 T; 8 other;
XX
XX Query Match 0.6%; Score 18; DB 21; Length 3529;
XX Best Local Similarity 100.0%; Pred. No. 1.5e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2831 aaagaacctctgaataa 2848
XX ||||||||||||||||
XX 3465 aaagaacctctgaataa 3482
XX
XX Db
XX
XX RESULT 71
XX AAV80604/C
XX ID AAV80604 standard; CDNA; 3942 BP.
XX
XX AAV80604;
XX
XX 01-MAR-1999 (first entry)
XX
XX Kidney injury associated molecule HM050 CDNA clone.
XX
XX Kidney injury associated molecule; kidney injury related molecule;
KW

KW KIM; tissue growth promotion; regeneration; renal condition;
KM acute renal failure; acute nephritis; tumour; ds.
XX
XX Rattus sp.
OS
XX
XX Key Location/Qualifiers
XX CDS 7..1329
XX
XX misc_feature /product= "kidney injury associated molecule"
XX complement (3542..3733)
XX /tag= b
XX /label= SAC_24644
XX
XX WO9853071-A1.
XX
XX 26-NOV-1998.
XX
XX 22-MAY-1998; 98WO-US10547.
XX
XX 23-MAY-1997; 97US-0047491.
XX 23-MAY-1997; 97US-0047490.
XX
XX (BIO) BIOGEN INC.
XX
XX Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
XX
XX WPI: 1999-045312/04.
XX P-PSDB; AAW86320.
XX
XX Kidney injury-associated molecule, KIM, polypeptides - upregulated
PT in injured or regenerating tissues, useful to promote tissue growth
PT and regeneration, especially to treat renal conditions
PT
XX Claim 9; Page 109-113; 213pp; English.
XX
XX The present sequence represents a kidney injury associated molecule
XX (KIM) CDNA clone. KIM proteins can be administered therapeutically
XX by expressing KIM encoding polynucleotides, to promote growth and/or
XX survival of damaged tissue (e.g. renal tissue), since the KIM proteins
XX are upregulated in injured or regenerating (especially renal) tissues.
XX KIM fusion proteins, conjugates, antibodies and vectors can also be used
XX therapeutically, e.g. these or the KIM proteins may be included with an
XX acceptable carrier in pharmaceutical compositions, useful for therapy/
XX prophylaxis of conditions associated with dysfunction/dysregulation of
XX KIM genes or proteins, especially renal diseases or impairments of renal
XX function in humans (e.g. acute renal failure, acute nephritis). The
XX polynucleotides can be used to produce antisense sequences which, when
XX internalised into cells, can disrupt expression of a cellular KIM gene,
XX also useful in therapy (e.g. to block the growth of tumours dependent on
XX KIM for growth) or compositions. The proteins and polynucleotides are
XX useful diagnostically e.g. to detect and quantify renal injury/disease
XX (indicative of increased risk, or presence of, renal injury or impaired
XX function), or abnormal responses to tissue injury (indicative of
XX increased risk, or presence of, an autoimmune response or abnormal
XX tissue growth arising from/affecting renal tissue). The proteins can
XX also be used to locate KIM-producing cells (especially specific loci,
XX e.g. tissue masses abnormally producing/expressing KIM such as tumours
XX arising from/affecting renal tissue), by contacting cells with an
XX imageable KIM-binding reagent and imaging reagent accumulation.
XX
XX Sequence 3942 BP; 908 A; 1019 C; 1087 G; 928 T; 0 other;
XX
XX Query Match 0.6%; Score 18; DB 20; Length 3942;
XX Best Local Similarity 100.0%; Pred. No. 1.5e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1918 cagaagacacagtgga 1935
XX ||||||||||||||||
XX 2495 cagaagacacagtgga 2478
XX
XX Db
XX
XX RESULT 72

ID	AA064473/c	standard; DNA: 3969 BP.
XX	AA064473;	
AC		
DT	02-JAN-2001	(first entry)
XX		
DE	DNA encoding an astaxanthin synthetase polypeptide.	
XX		
KW	Astaxanthin synthetase; astaxanthin; beta-carotene; carotenogenic yeast; antioxidant; cancer; colouring reagent; farmed fish; salmon; ss.	
XX		
OS	Phaffia rhodozyma.	
XX		
FH	Key	Location/Qualifiers
FJ	CDS	535..3700
FJ		/*tag= a
FJ		/product= "astaxanthin synthetase"
FJ	5'UTR	517..518
FJ		/*tag= b
FJ	exon	535..652
FJ		/*tag= c
FJ		/number= 1
FJ	intron	653..734
FJ		/*tag= d
FJ		/number= 1
FJ	exon	735..783
FJ		/*tag= e
FJ		/number= 2
FJ	intron	784..898
FJ		/*tag= f
FJ		/number= 2
FJ	exon	899..1015
FJ		/*tag= g
FJ		/number= 3
FJ	intron	1016..1087
FJ		/*tag= h
FJ		/number= 3
FJ	exon	1088..1179
FJ		/*tag= i
FJ		/number= 4
FJ	intron	1180..1302
FJ		/*tag= j
FJ		/number= 4
FJ	exon	1303..1517
FJ		/*tag= k
FJ		/number= 5
FJ	intron	1518..1600
FJ		/*tag= l
FJ		/number= 5
FJ	exon	1601..1634
FJ		/*tag= m
FJ		/number= 6
FJ	intron	1635..1723
FJ		/*tag= n
FJ	exon	1724..1866
FJ		/*tag= o
FJ		/number= 7
FJ	intron	1867..1939
FJ		/*tag= p
FJ		/number= 7
FJ	exon	1940..1999
FJ		/*tag= q
FJ		/number= 8
FJ	intron	2000..2081
FJ		/*tag= r
FJ		/number= 8
FJ	exon	2082..2181
FJ		/*tag= s
FJ		/number= 9
FJ	intron	2182..2257
FJ		/*tag= t
FJ		/number= 9

FT	exon	2258..2354
FT		/*tag= u
FT		/number= 10
FT	intron	2355..2431
FT		/*tag= v
FT		/number= 10
FT	exon	2432..2542
FT		/*tag= w
FT		/number= 11
FT	intron	2543..2618
FT		/*tag= x
FT		/number= 11
FT	exon	2619..2652
FT		/*tag= y
FT		/number= 12
FT	intron	2653..2742
FT		/*tag= z
FT		/number= 12
FT	exon	2743..2814
FT		/*tag= aa
FT		/number= 13
FT	intron	2815..2962
FT		/*tag= ab
FT		/number= 13
FT	exon	2963..3050
FT		/*tag= ac
FT		/number= 14
FT	intron	3051..3113
FT		/*tag= ad
FT		/number= 14
FT	exon	3114..3171
FT		/*tag= ae
FT		/number= 15
FT	intron	3172..3247
FT		/*tag= af
FT		/number= 15
FT	exon	3248..3321
FT		/*tag= ag
FT		/number= 16
FT	intron	3322..3398
FT		/*tag= ah
FT		/number= 16
FT	exon	3399..3423
FT		/*tag= ai
FT		/number= 17
FT	intron	3424..3513
FT		/*tag= aj
FT		/number= 17
FT	exon	3514..3700
FT		/*tag= ak
FT		/number= 18
FT	polyA_signal	3865..3866
FT		/*tag= al
XX	EP1035206-A1.	
XX		
XX	13-SEP-2000.	
XX		
XX	03-MAR-2000; 2000EP-0104430.	
XX		
XX	09-MAR-1999; 99EP-0104668.	
XX	01-FEB-2000; 2000EP-0101666.	
XX		
XX	(HOFF) HOFFMANN LA ROCHE & CO AG F.	
XX		
XX	Hoshino T, Ojima K, Setoguchi Y;	
XX		
XX	WPI; 2000-559874/52.	
XX	P-PSDB; AAB08713.	
XX		
XX	Novel polynucleotide encoding astaxanthin synthase useful for producing recombinant cells for producing astaxanthin from beta-carotene	
XX		

PS Claim 4; Page 30-33; 46pp; English.

XX The present sequence encodes an astaxanthin synthetase polypeptide

CC of Phaffia rhodozyma. The enzyme is involved in the last step of the

CC astaxanthin biosynthesis pathway, from beta-carotene to astaxanthin.

CC P. rhodozyma is a carotenogenic yeast strain. The astaxanthin

CC synthetase polynucleotides and polypeptides are useful for producing

CC astaxanthin. Astaxanthin is an antioxidant which may be used to

CC protect living cells against diseases such as cancer. Astaxanthin is

CC also used as a colouring reagent, e.g. in farmed fish like salmon to

CC impart an orange-red coloration.

XX

SO Sequence 3969 BP; 937 A; 942 C; 869 G; 1221 T; 0 other;

Query Match 0.6%; Score 18; DB 21; Length 3969;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 420 aactgaaagaccctc 437

|||||

Db 1193 AACTTGAAAGCACCTTC 1176

RESULT 73

AA158415 standard; cDNA: 4215 BP.

AA158415;

XX

XX 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 618.

XX

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokine; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX

XX Homo sapiens.

OS

XX

XX WO200153312-A1.

PN

XX 26-JUL-2001.

PD

XX

XX 26-DEC-2000; 2000WO-US34263.

PF

XX

XX 21-JAN-2000; 2000US-0488725.

PR

XX 25-APR-2000; 2000US-0552317.

PR

XX 09-JUL-2000; 2000US-0598042.

PR

XX 19-JUL-2000; 2000US-0620312.

PR

XX 03-AUG-2000; 2000US-0653450.

PR

XX 14-SEP-2000; 2000US-0662191.

PR

XX 19-OCT-2000; 2000US-0693036.

PR

XX 29-NOV-2000; 2000US-0727344.

PR

XX (HYSEQ-) HYSEQ INC.

PA

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX

XX WPI: 2001-442253/47.

DR

XX P-PSDB; AAM39259.

XX

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX

XX Claim 1; SEQ ID NO 618; 10078pp; English.

XX

XX The invention relates to human nucleic acids (AA157798-AA161365) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.

XX

SO Sequence 4215 BP; 1523 A; 901 C; 855 G; 934 T; 2 other;

Query Match 0.6%; Score 18; DB 22; Length 4215;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2831 aaagaccctctgaaaaa 2848

|||||

Db 3248 aaagaccctctgaaaaa 3265

RESULT 74

AA160201 standard; cDNA: 4871 BP.

AA160201;

XX

XX 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 4190.

XX

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokine; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX

XX Homo sapiens.

OS

XX

XX WO200153312-A1.

PN

XX 26-JUL-2001.

PD

XX

XX 26-DEC-2000; 2000WO-US34263.

PF

XX

XX 21-JAN-2000; 2000US-0488725.

PR

XX 25-APR-2000; 2000US-0552317.

PR

XX 09-JUL-2000; 2000US-0598042.

PR

XX 19-JUL-2000; 2000US-0620312.

PR

XX 03-AUG-2000; 2000US-0653450.

PR

XX 14-SEP-2000; 2000US-0662191.

PR

XX 19-OCT-2000; 2000US-0693036.

PR

XX 29-NOV-2000; 2000US-0727344.

PR

XX (HYSEQ-) HYSEQ INC.

PA

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX

XX WPI: 2001-442253/47.

DR

XX P-PSDB; AAM41045.

XX

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4190; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA138642-AA142213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SO Sequence 4871 BP; 1815 A; 1021 C; 990 G; 1045 T; 0 other;
 XX
 XX
 Query Match 0.6%; Score 18; DB 22; Length 4871;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2831 aaagaacctctgaaaaa 2848
 ||||||||||||||||
 DB 3258 aaagaacctctgaaaaa 3275
 RESULT 75
 AA13198/C
 ID AA13198 standard; DNA: 4951 BP.
 XX
 AC AA13198;
 XX
 DT 19-MAR-1999 (first entry)
 XX
 DE Enterococcus faecalis genome contig SEQ ID NO:261.
 XX
 KM Enterococcus faecalis; contig; detection; Enterococcal infection;
 KM vaccine; attenuation; computer readable medium; ds.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO9850555-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98MO-US08985.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Dillon PJ, Kunsch CA;
 XX
 DR WPI; 1999-045171/04.
 XX
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 XX
 PS Claim 1; Page 1248-1251; 2084pp; English.
 CC
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AA12938 to AA13919 represent these nucleotide sequences which are

CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 CC
 SO Sequence 4951 BP; 1563 A; 981 C; 735 G; 1665 T; 7 other;
 XX
 XX
 Query Match 0.6%; Score 18; DB 20; Length 4951;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2527 gcttgctgcaaatgcag 2544
 ||||||||||||||||
 DB 1288 GCTTGCTGCAAAATGCAG 1271
 Search completed: March 25, 2002, 16:50:28
 Job time: 9092 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:40:31 ; Search time 2764.97 Seconds
(without alignments)
12176.090 Million cell updates/sec

Title: US-09-697-089-1

Perfect score: 3133
Sequence: 1 cgcctcgcgcgcggtgggaag.....aagtgctacgaagccacta 3133

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 537288281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gp_estcl:*
11: gp_estl2:*
12: gp_hic:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	551	17.6	552	13	AO309404 CITBT-EL-
C 2	446	14.2	480	10	AV719179
C 3	364	11.6	364	10	AI263294 qx57b01.x
C 4	291	9.3	371	10	AV656315 AV656315
C 5	249	7.9	261	10	AM337918 he12h11.x
C 6	172	5.5	476	11	BG210375 RST29913
C 7	170	5.4	553	13	AO320928 RPTC11-93
C 8	162	5.2	509	10	AM418826 ha21e11.x
C 9	146	4.7	630	13	AQ112439 CIT-HSP-2
C 10	124	4.0	412	11	H25984 Y156907.t1
C 11	106	3.4	470	13	AO624020 HS-5378.B
C 12	93	3.0	499	10	AI023795 OX08403.x

C 13	81	2.6	840	11	BF207840
C 14	58	1.9	219	13	AO283886
C 15	55	1.8	635	10	B58691
C 16	32	1.0	446	10	AI222422
C 17	30	1.0	251	11	BF903662
C 18	28	0.9	404	13	AO889169
C 19	25	0.8	443	13	AO497022
C 20	25	0.8	674	13	AO541295
C 21	23	0.7	606	11	BF829853
C 22	22	0.7	256	10	AV268403
C 23	22	0.7	437	13	A2740007
C 24	21	0.7	163	13	A2086396
C 25	21	0.7	343	10	AA784099
C 26	21	0.7	343	10	AA821339
C 27	21	0.7	371	10	AI645009
C 28	21	0.7	432	13	A2223093
C 29	21	0.7	437	13	AO555443
C 30	21	0.7	462	13	CNS00592
C 31	21	0.7	602	13	A2360053
C 32	21	0.7	1147	11	BG571228
C 33	20	0.6	64	11	BI142462
C 34	20	0.6	73	10	BE638399
C 35	20	0.6	205	13	AO069062
C 36	20	0.6	215	11	BI315256
C 37	20	0.6	216	10	AA493806
C 38	20	0.6	225	10	AI366637
C 39	20	0.6	248	10	AI603832
C 40	20	0.6	299	13	A2891870
C 41	20	0.6	332	10	AM357412
C 42	20	0.6	336	13	AO952434
C 43	20	0.6	342	10	AA670222
C 44	20	0.6	380	13	A2991565
C 45	20	0.6	333	10	AA294641
C 46	20	0.6	333	10	AM838955
C 47	20	0.6	415	10	AA294540
C 48	20	0.6	415	11	T81898
C 49	20	0.6	439	10	AA294143
C 50	20	0.6	483	13	AO209355
C 51	20	0.6	481	11	BF601894
C 52	20	0.6	498	10	AI643174
C 53	20	0.6	505	11	BF602580
C 54	20	0.6	519	11	BG382487
C 55	20	0.6	521	10	AI670486
C 56	20	0.6	521	13	FR0031506
C 57	20	0.6	526	11	BI394563
C 58	20	0.6	535	11	BG351255
C 59	20	0.6	541	13	A2221490
C 60	20	0.6	554	13	A2298434
C 61	20	0.6	557	13	AO270247
C 62	20	0.6	571	11	BF258009
C 63	20	0.6	577	11	BG385206
C 64	20	0.6	577	13	AO570315
C 65	20	0.6	580	13	A2987277
C 66	20	0.6	605	11	BG098273
C 67	20	0.6	620	13	A2802117
C 68	20	0.6	627	10	AI511575
C 69	20	0.6	633	13	A2510047
C 70	20	0.6	644	10	AM874886
C 71	20	0.6	652	10	BE346038
C 72	20	0.6	653	10	AI317901
C 73	20	0.6	656	10	BE636495
C 74	20	0.6	660	11	BG310513
C 75	20	0.6	726	13	CNS02935
C 76	20	0.6	729	13	B65395
C 77	20	0.6	731	11	BG250425
C 78	20	0.6	754	13	AO329189
C 79	20	0.6	760	10	AI591096
C 80	20	0.6	772	11	BG310534
C 81	20	0.6	779	11	BG350968
C 82	20	0.6	783	11	BI184833
C 83	20	0.6	784	11	BF824698
C 84	20	0.6	891	11	BG248945
C 85	20	0.6	937	13	CNS02507

c	86	20	0.6	939	13	CNS02105	AL176270 tetradodon
c	87	20	0.6	1036	11	BG176063	BG176063 602335288
c	88	20	0.6	1396	11	BG254706	BG254706 602368693
c	88	19	0.6	207	10	AV656506	AV656506 AV656506
c	50	19	0.6	209	10	AV144358	AV144358 AV144358
c	92	19	0.6	238	10	AV371617	AV371617 AV371617
c	91	19	0.6	256	10	AI340263	AI340263 9K86601_x
c	93	19	0.6	272	11	BG381720	BG381720 UT-R-CTO-
c	94	19	0.6	283	10	AV048726	AV048726 AV048726
c	95	19	0.6	304	13	AQ094693	AQ094693 HS_3025_B
c	96	19	0.6	314	13	B62990	B62990 CIT-HSP-714
c	97	19	0.6	322	10	BB318669	BB318669 BB318669
c	98	19	0.6	334	13	AQ285454	AQ285454 RRCIT1-90
c	99	19	0.6	351	13	AZ784848	AZ784848 RRCIT-24-1
c	100	19	0.6	353	10	AW356070	AW356070 36057 MAN

ALIGNMENTS

[illegible]

ORGANISM	TITLE
Homo sapiens	Use of a random human BAC End Sequence Database for Sequence-Ready
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
1 (Bases 1 to 552)	
ADAMS, M.D., GOUNSELEY, S.D., ZHAO, S., BASS, S., LINHER, K., GOLDEN, K.,	
BERRY, K.K., RANNEY, D., SHIH, E., WIBLE, C., SHIZUYA, H., SIMON, M.,	
VENTER, J.C.	

JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CITBI-EI-2528J13.TR

Contact: Shanying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeet@igr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .552

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2528313"
/clone_id="CITB1-E1"
/sex="male"
/cell_type="sperm"
/notes="Vector: pBelobAC1; Site_1: EcoRI; Site_2: EcoRI
CalTech Human BAC Library D"
BASE COUNT
144 a 121 c 130 g 157 t

```

```

Query Match      17.6%   Score 551:  DB 13: length 552:
Best Local Similarity 100.0%:  Pred. No. 6.6e-267:
Matches 551; Conservative 0; Mismatches 0; Indels 0

```

OY		1462	aaatggttcattcgcagacattaaccacttatatgcagctgctccgyltacaccttg	1528
Db		552	AATGCTTCCATTTCGACATTACATCCACTTATAGCAGCGCTGCCGGTTACACCCTTG	493

Oy	1522	ggaatcctgaggaaagccaaagggcgtgataaggaagcttcgacaagatgataaacg	1581
Db	432	ggatcatctgttgaaagccaaacccaggcgtttatgaaagacctccgagatgtaacacg	433
Oy	1582	gcgcctctctcgaacttcacatgcgaagaagcctctctgagadaggaactcttgcaaa	1641
Db	432	gctccctcttcgcgactttccatcgccaaagagccctctctgagacaggaactcttgcaaa	373
Oy	1642	gttgaaaaaacacacatgagcaagaactctctgaaagccataagctaatctcttgag	1701
Db	372	gttgaaaaaacacacatgagcaagaactctctgaaagccataagctaatctcttgag	313
Oy	1702	agtgatgcattcatatatacaagagagtaacatccaaatcaagcccttgagccaagaattg	1761
Db	312	agtgatgcattcatatatacaagagagtaacatccaaatcaagcccttgagccaagaattg	253
Oy	1762	aagcttctcttcaagatgaaaagcttctatatacaactcagaagaaactcccgattactat	1821
Db	252	aagcttctcttcaagatgaaaagcttctatatacaactcagaagaaactcccgattactat	193
Oy	1822	ctgactctcttgaaacatttgcaccaatctgagcaatgctcttgactctataaactgagct	1881
Db	192	ctgactctcttgaaacatttgcaccaatgagcaatgctcttgactctataaactgagct	133
Oy	1882	cttctgaggagagctatgagcttcaatggaaaagctgcacaaagacaaagttgaaatccaca	1941
Db	132	cttctgaggagagagctatgagcttcaatggaaaagctgcacaaagacaaagttgaaatccaca	73
Oy	1942	tggaagaagcccccaaaaacattacatctccagacagagctgtaactttgtctctcaactgga	2001
Db	72	tggaagaagcccccaaaaacattacatctccagacagagctgtaactttgtctctcaactgga	13
Oy	2002	agcagaatc 2012	
Db	12	agcagaatc 2	

RESULT	2							
AV719179/c								
LOCUS	AV719179	480 bp	mRNA	EST	16-OCT-2000			
DEFINITION	AV719179	GLC Homo sapiens	CDNA	clone	GLCEAD10.5',	mRNA	sequence	
ACCESSION	AV719179							
VERSION	AV719179.1	GI:10816331						
KEYWORDS	EST.							
SOURCE	human.							

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 480)
AUTHORS
Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,
Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,

TITLE	Homo sapiens cDNA GLC clones
JOURNAL	Unpublished (2000)
COMMENT	Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzgc@ncgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1..480

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GIC60A10"
/clone_lib="GIC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"

```


/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 138 a 120 c 85 g 137 t
 ORIGIN

Query Match 14.2%; Score 446; DB 10; Length 480;
 Best Local Similarity 100.0%; Pred. No. 7.4e-214;
 Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2640 catgaactgacagacagatgaacgtgctagaacagctcaccgactgactgctccctg 2699
 |||||||
 DB 480 CATGAACCTGATCGACAGATGAACGCTAGAAACACCTCACCTGATGCTGCCCTG 421
 |||||||
 QY 2700 gctctggaagctgaagcagcctgagcagcctgcttgaacatttgagagaggtcccaaa 2759
 |||||||
 DB 420 GGGTGTGACGTGCAAGCAGCCCTGAGACCCCTGTTAAACATTGAGAGGAGGCCACAA 361
 |||||||
 QY 2760 ctcgtcaagcttgggttgaacaaactgagacacacagatagaattagaattagtt 2819
 |||||||
 DB 360 CTCGTCAACCTTGGCTTGAACAACTGAGACCTCACAGATACAGATTAGATTAGT 301
 |||||||
 QY 2820 gcaatttttgaagaagaccctctgaacaaactccagcagttgaatttggcggaatcgt 2879
 |||||||
 DB 300 GCATTTTGGAAAGAACCCCTCGAAAAAATTCCACAGATTGAATTGGCGGGAATCGT 241
 |||||||
 QY 2880 gtgagcagtgatgagtgctgctcctcatagtggtgtatttgagaatttaagcattagt 2939
 |||||||
 DB 240 GTGAGCAGTATGATGAGCTTGCTTCATGAGGTGTATTGAGAACTTAAACATTAGTG 181
 |||||||
 QY 2940 tttttgactttagtactaagaatttctacactgacacagcattagtcagaacaaactagc 2999
 |||||||
 DB 180 TTTTGTGACTTATGATCTAAGAAATTTCTACCGATCCAGCATTAATGCAAAAACTTAGC 121
 |||||||
 QY 3000 caagtgtatccaagtttaacttctctgcagaagcagcagctgttgggtggcaattgat 3059
 |||||||
 DB 120 CAAGTGTATTCCAAGTTAACTTTTCGCAAGAGCTAGGCTTGTGGTGGCAATTTGAT 61
 |||||||
 QY 3060 gatgatgactcagtgatttactagc 3085
 |||||||
 DB 60 GATGATGATCTCAGTGTATTACAG 35

RESULT 3
 LOCUS A1263294 364 bp mRNA EST 03-FEB-1999
 DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
 mRNA sequence.
 ACCESSION A1263294
 VERSION A1263294.1 GI:3871497
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 364)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
 www.bio.lnl.gov/bbrp/image/image.html
 Insert Length: 2146 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 364.
 Location/Qualifiers
 1. 364
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="IMAGE:2005417"
 /clone_lib="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: PCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"
 BASE COUNT 117 a 84 c 55 g 108 t
 ORIGIN

Query Match 11.6%; Score 364; DB 10; Length 364;
 Best Local Similarity 100.0%; Pred. No. 2e-172;
 Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2766 aagcttgggttgaacaaactgagacacacagatagaattagaatttagtgcatt 2825
 |||||||
 DB 364 AACCTTGGGTTGAACAACTGAGACCTCACAGATACAGATTAGAAATTTAGGTGCATT 305
 |||||||
 QY 2826 ttggaagaagaccctctgaacaaactccagcagttgaatttggcggaatcgtgtgagc 2885
 |||||||
 DB 304 TTGGAAAGAACCCCTCGAAAAAATTCCACAGATTGAATTGGCGGGAATCGTGTGAGC 245
 |||||||
 QY 2886 agtgatgagtgctgctcctcatagtggtgtatttgagaatttaagcattagtgtttt 2945
 |||||||
 DB 244 AGTGATGATGATGCTTGCTTCATGAGGTGTATTGAGAACTTAAACATTAGTGT 185
 |||||||
 QY 2946 gactttagtactaagaatttctacactgacacagcattagtcagaacaaactagc 3005
 |||||||
 DB 184 GACTTATGATCTAAGAAATTTCTACCGATCCAGCATTAATGCAAAAACTTAGCAGT 125
 |||||||
 QY 3006 ttatccaagtttaacttctctgcagaagcagcagctgttgggtggcaatttgatgat 3055
 |||||||
 DB 124 TTATCCAAAGTTAACTTTTCGCAAGAGCTAGGCTTGTGGTGGCAATTTGATGATGAT 65
 |||||||
 QY 3066 gatcagtgatttactagtgctttaaactgtaactgtaactgtaactgtaactgtaactg 3125
 |||||||
 DB 64 GATCTCAGTGTATTACAGTGCTTTTAACTAGTAACTGTTAAATTAAGTACTCGA 5
 |||||||
 QY 3126 agcc 3129
 |||||
 DB 4 AGCC 1

RESULT 4
 LOCUS AV656315 371 bp mRNA EST 07-SEP-2000
 DEFINITION AV656315 GUC Homo sapiens cDNA clone G1CEGA10 3', mRNA sequence.
 ACCESSION AV656315
 VERSION AV656315.1 GI:9877329
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 371)
 REFERENCE Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
 Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
 Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
 G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
 Homo sapiens cDNA clone
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@cnhc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers

FEATURES
 source 1. 364
 /organism="Homo sapiens"

FEATURES
 source 1. 364
 /organism="Homo sapiens"

source 1..371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCEQA10"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 112 a 85 c 91 g 82 t 1 others
ORIGIN

Query Match 9.3%; Score 291; DB 10; Length 371;
Best Local Similarity 99.7%; Pred. No. 1.5e-135;
Matches 341; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1908 gaaagctgcaagaacacagcgtggaatccacatggaagagcccaaacctacatt 1967
|||
Db 1 GAAAGGCTGAGAGACACAGGTGGATCCACATGGAAGGCCCCAGAAACTACAT 60
|||
QY 1968 cccagcagcgtgctatcttcttctcaactggaagcaggaattcaggacctggaagtc 2027
|||
Db 61 CCCAGCAGGCTGTRATCTTTGTTCTTCACTGGAAGCAGAAATTCAGACTGTG 120
|||
QY 2028 aacatccggaattcagaagttgataagaagatcacatcctgagggaatattc 2087
|||
Db 121 ACATCCCGGATTCAGCAAGTTGAATAGCAAGATTCAGATATCTGGGAAATATTC 180
|||
QY 2088 agctctgcacaaagcctcagcgtgcaaaataagagatgctgctgctggaagcctc 2147
|||
Db 181 AGCTCTGCCAAGAGCTCAGCTGCAGAAATGAAGATGTGCTGCTGGAAGCCTC 240
|||
QY 2148 agttggtccctcagacactgaaagacattatctcctcagtggaagcagtcctc 2207
|||
Db 241 AGTTGGCTCCTCAGCAGCTGTGAAGACATTTATCTCTCATGAGGGAAGCCAGTCCCTC 300
|||
QY 2208 accatagaagatggaagacatcacatctgtaacaaactg 2249
|||
Db 301 ACCATAGAAAGATGAGAGCAGCATCATCTGTAAACAACCTG 342
|||

RESULT 5
AW337918/c 261 bp mRNA EST 31-JAN-2000
LOCUS he12h1.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2918853 3'
DEFINITION mRNA sequence.
ACCESSION AW337918
VERSION AW337918.1 GI:6834544
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 261)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: -40up from Gibco
High quality sequence stop: 201.
Location/Qualifiers

FEATURES

source 1..261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2918853"
/clone_lib="NCI_CGAP_CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: pCMV-SPORT6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Library constructed by Life Technologies."

BASE COUNT 86 a 54 c 38 g 83 t
ORIGIN

Query Match 7.9%; Score 249; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.5e-114;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2885 cagtatgagatgctgctcctcatggtgatttgaagatcctaagcaattagtgttt 2944
|||
Db 261 CAGTATGATGATGCTGCTCCCTTCATGGTGTATTGAGAACTTAAGCAATTAGTGT 202
|||
QY 2945 tgactttagtaagaatttctacacctgataccagcattagtcagaagaactagccaagt 3004
|||
Db 201 TGACTTTAGTACTAAGAATTCTCTACCTGATCCAGCATTAAGTCAGAAACTAGCCAGT 142
|||
QY 3005 gtatccaaagttaactttctgcagaagaagctagctgtgttggtgcgaatttgatgaga 3064
|||
Db 141 GTTATCAAGTTAACTTTCTGCAAGAGCTAGGCTTGTGGGTGCAATTGTGATGATGA 82
|||
QY 3065 tgactcagtggttactaacagtggtctttaaactagtaactgcttaataaggttactgc 3124
|||
Db 81 TGATCTCATGCTGTTATTACAGAGTCTTTTAACCTAGTAACCTGTTAAATTAAGTACTCG 22
|||
QY 3125 aagccagta 3133
|||
Db 21 AAGCCAGTA 13
|||

RESULT 6
BG210375 476 bp mRNA EST 21-APR-2001
LOCUS RST29913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION RST29913
ACCESSION BG210375
VERSION BG210375.1 GI:13732062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 476)
Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Krasnoc, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
Velasco, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
and Ducar, M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 360.
Location/Qualifiers

FEATURES

	/cell_line="HT1080"
	/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, In press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."
BASE COUNT	120 a 99 c 107 g 149 t 1 others
ORIGIN	
Query Match	5.5% Score 172; DB 11; Length 476;
Best Local Similarity	99.6%; Pred. No. 2.1e-75;
Matches 222:	Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Y 2908	tgggtgatattgagaatcttaagcaatcagtcgttttctaacttagttaagaatttc 2967
Db 242	TGGGCTATTTCAGCAATCTTAAGCAATTAGTGTTTTTCACCTTTAGTACTAAAGAATTTCC 301
Y 2968	tacctgacgcagcatctagtcagaaaaacttagccaagtgtlccaaagtaactttctgc 3027
Db 302	TACCGATGCCACCATTTAGTCAGAAAACCTTAGCCAGTGTTATTCCAAAGTTAATCTTCTGC 361
Y 3028	aagaagctcagcgtctgttggtgagcaatttgatgatgatgaatccagctgttatccaggtg 3087
Db 362	AAGAAGCTAAGGCTTGTGGGTGGCAATTTGATGATGATGATCTCAGTGCCTATTACAGGTG 421
Y 3088	cctttaaacgtlaactgctttaataaaggtaagtcgaagcca 3130
Db 422	CTTTTAAACTAGTAACTGCTTTAAATAAAGTGATACGAAGCCA 464

RESULT	7
LOCUS	AQ320928/c
DEFINITION	AO320928 553 bp DNA GSS 06-MAY-1999
ACCESSION	RPC111-93C9.TV RPC1-11 Homo sapiens genomic clone RPC1-11-93C9, DNA sequence.
VERSION	AQ320928
KEYWORDS	AQ320928.1 GI:4053662
SOURCE	GSS.
ORGANISM	human.
EREFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 553)
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P., and Venter,J.C. Use of human BAC End Sequences for Sequence-Ready Map Building Unpublished (1998)
TITLE	Other,GSS: RPC111-93C9.TI
JOURNAL	Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208
COMMENT	Email: bheetjgr.org Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong [pieter@dejong.med.buffalo.edu]. Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html Seq primer: t7 Class: BAC ends
FEATURES	Location/Qualifiers
source	1..553

```
FEATURES
source      location/Qualifiers
            1..553
/organism="Homo sapiens"
/db_xref="GDB:7535384"
/db_xref="taxon:9606"
/clone="PC1-11-93C9"
/clone_1b="PC1-11"
/sex="Male"
```

	Query Match	Best Local Similarity	Matches 170: Conservative	5.4%: Pred. NO. 2.1e-74	Score 170: DB 13: Length 553;	0; Indels 0; Gaps 0;
Oy	2650	tcgaacagatgaacgtgctactgaacagcctcaccgcactgtgtcccttgaggctgtgacg	100.0%	0;	2709	
Db	258	TCGACAGGATGTAAGCTGCTAGAACAGCTCACCGCACCTGATGCTGCCCTTGGGGCTGTGAGC	0;	0;	199	
Oy	2710	tgcaagcagcctctgagcagcctgttgaacaatttggaggaagtlcccaacaactcgtlcaagc	100.0%	0;	2769	
Db	198	TGCAAGGCAGAGCTGTGAGCAGCGCTGTGAAACAATTTGGAGAGAGGTCCACCAACTGCTCAAC	0;	0;	139	
Oy	2770	ttgggttgaaaaactgtgagacatcacacagatlacagagatltgaatttgaatt	100.0%	0;	2819	
Db	138	TTGGGTTGAAAACACTGAGACTCACAGATPACAGAGATTGAAATTTTAGGT	0;	0;	89	

FEATURES	source
RESULT 8	
LOCUS	AM418826/c
DEFINITION	AM418826 509 bp mRNA EST 09-FEB-2000
ACCESSION	ha21el1.x1 NCI_CGAP_Kid12 Homo sapiens CDNA clone IMAGE:2874380 3',
VERSION	AM418826
KEYWORDS	mRNA sequence.
SOURCE	AM418826.1 GI:6946758
ORGANISM	EST.
	human.
REFERENCE	human, sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (bases 1 to 509)
COMMENT	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
	Tumor Gene Index
	Unpublished (1997)
	Contact: Robert Strausberg, Ph.D.
	Email: cgapbs-remail.nih.gov
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
	Emmert-Buck, M.D., Ph.D.
	CDNA Library Preparation: M. Bento Soares, Ph.D.
	CDNA Library Arrayed by: Greg Lennon, Ph.D.
	DNA Sequencing by: Washington University Genome Sequencing Center
	clone distribution: NCI-CGAP clone distribution information can be
	found through the I.M.A.G.E. Consortium/TLNL at:
	www-bio.linnl.gov/bdrrp/image/image.html
	Seq primer: -40up from G1bco
	High quality sequence stop: 461.
	Location/Qualifiers
	1..509

FEATURES	SOURCE
Location/Qualifiers	1..509
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:2874380"	
/clone_lib="NCI-CGAP_Ki67"	
/tissue_type="2 pooled tumors (clear cell type)"	
/lab_host="DH10B"	
/note="Organ: Kidney; Vector: pT7A3-Pac (pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP_Ki67 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs:1329912-1325831, 1471366-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	161 a 99 c 80 g 169 t
ORIGIN	

Query Match 5.2%; Score 162; DB 10; Length 509;
Best Local Similarity 100.0%; Pred. No. 2.3e-70;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2972 tgalccagcattatgcagaaaactcagcagatgtatccaaagtttaactttctcaga 3031
|||||
Db 509 TGATCCAGCATTAGTACAGAAAACCTTAGCCAGTGTATTCAGGTTAACTTTCTGCAGAA 450

Qy 3032 agctagcttctgtggtgcaatttgatgatgatcctcagtttatcagaggtctt 3091
|||||
Db 449 AGCTAGAGCTTGTGGTGGTGCATTTGATGATGATCATCTCACTGTTATTACAGGTCTT 390

Qy 3092 taactagtaactgcttaataaagtgactcgaagccagta 3133
|||||
Db 389 TAACTAGTAACTGCTTAATAAAGTGTACTGAGAGCCAGTA 348

RESULT 9
A0112439 630 bp DNA GSS 29-AUG-1998
LOCUS CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA
DEFINITION sequence.
ACCESSION A0112439
VERSION A0112439.1 GI:3484599
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 630)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Sub,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other-GSSs: CIT-HSP-2372C1.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
Source location/Qualifiers
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2372C1"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC1; site_1: HindIII; site_2:
HindIII"

BASE COUNT 188 a 126 c 124 g 192 t
ORIGIN

Query Match 4.7%; Score 146; DB 13; Length 630;
Best Local Similarity 100.0%; Pred. No. 2.9e-62;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2384 agctgaagcctgtaaaacctgaagaagatgtttatcattgaccactgtctga 2443
|||||
Db 485 AGCTGAAGCCTGTAAGAAACCTGAAGAAGATGTGTTATTATTGACCCACTGTCTGA 544

Qy 2444 catgagagagggaatgattacatagtcgaagtcctctgtcgaagtgaacctgtgacctga 2503
|||||
Db 545 CATGAGAGAGGGAATGATTACATAGTCAAGTCTGTGCAAGTGAACCTGTGACCTTGA 604

Qy 2504 agaattcaattagctctctgtctgt 2529
|||||
Db 605 AGAAATTCATTAATGATGCTCTGCTGCT 630

RESULT 10
H25984/c 412 bp mRNA EST 10-JUL-1995
LOCUS Y156907.r1 Soares Breast 3NbHst Homo sapiens cDNA clone
DEFINITION IMAGE:162300 5', mRNA sequence.
ACCESSION H25984
VERSION H25984.1 GI:895107
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 412)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maitra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE The Washu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 942
High quality sequence stops: 347
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 942 Std Error: 0.00
Seq primer: M13Rpl
High quality sequence stop: 347.
FEATURES
Source location/Qualifiers
1..412
/organism="Homo sapiens"
/db_xref="GDB:576544"
/db_xref="taxon:9606"
/clone="IMAGE:162300"
/clone_lib="Soares Breast 3NbHst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAAGTGGAGCGGCCCTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of a modified pT73 vector (Pharmacia).
Library went through one round of normalization to a Cot =
20. Library constructed by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 129 a 84 c 73 g 124 t 2 others
ORIGIN

Query Match 4.0%; Score 124; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 3.7e-51;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3010 ccaagttacttctctgaaagaagctagctgttgggtgcaattgatgatgatc 3069
|||||
Db 340 CCAAGTTACTTCTCTGAAAGAGCTAGCTGTGGGTGCAATTGATGATGATGATC 281

Db 3070 tcagtgatatacaggctgtcttaactaaactgaactgaatgaagctgaagcc 3129
|||||
Db 280 TCAGTGTATATACAGGCTTTTAACACTAGCTTAATAAAGTACTGCGAAGCC 221
Qy 3130 agta 3133
|||||
Db 220 AGTA 217

RESULT 11
AO624020 470 bp DNA GSS 16-JUN-1999
LOCUS HS.5378.B2.C12_SPE RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=954 Col=24 Row=F, DNA sequence.
ACCESSION AO624020
VERSION AO624020.1 GI:5086412
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 470)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@edj.med.buffalo.edu). Clones may be purchased from
BACpac Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web server:
http://www.htsc.washington.edu
Plate: 954 row: F column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 470.
FEATURES
source
Location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=954 Col=24 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"

BASE COUNT 142 a 101 c 107 g 117 t 3 others

ORIGIN
Query Match 3.4%; Score 106; DB 13; Length 470;
Best Local Similarity 100.0%; Pred. No. 4.6e-42;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2069 atactcgggaaataatcagctcgcacaaagccacagctcacaataaagagatgac 2128
|||||
Db 64 ATTATCGGGGAAATATTCAGCTCTGCCACAAAGCTCAGGCTCAATTAAGAGATGTGC 123
Qy 2129 tggctgtgctggaagcctcagtttgctcagcacctgtaagaac 2174
|||||

Db 124 TGGTGTGCTGGAAGCCTCAGTTTGTGCTCAGACCTGTAGAAC 169
RESULT 12
AI023795 499 bp mRNA EST 28-AUG-1998
LOCUS ox08d03.x1 Soares_fetal_liver_spleen.INFIS.S1 Homo sapiens cDNA
DEFINITION clone IMAGE:1655717 3', mRNA sequence.
ACCESSION AI023795
VERSION AI023795.1 GI:3238839
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 499)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1179 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 439.
FEATURES
source
Location/Qualifiers
1..499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1655717"
/clone_lib="Soares_fetal_liver_spleen.INFIS.S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFIS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AAGTCGACGATTAATTAAGACCTTTGCTTGTGCTGAGATTTAAAGGACT 3'],
AAGTCGACGATTAATTAAGACCTTTGCTTGTGCTGAGATTTAAAGGACT (Pharmacia),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 144 a 109 c 98 g 148 t

ORIGIN
Query Match 3.0%; Score 93; DB 10; Length 499;
Best Local Similarity 100.0%; Pred. No. 1.7e-35;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 gttctttatcagaataagaagagacttgacattgtctcggattaaagact 357
|||||
Db 407 GTCTTTTTCATCAGACATCGAAGAGACTTGACGATTTGCTGAGATTTAAAGGACT 466
Qy 358 tttacatacccatcttctgaactttatc 390
|||||
Db 467 TGTACCAFAACCCCATCTTTCTGACACTTTATC 499

RESULT 13
BF207840 840 bp mRNA EST 06-NOV-2000
LOCUS BF207840/c
DEFINITION 601862546f1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082074 5',
mRNA sequence.
ACCESSION BF207840
VERSION BF207840.1 GI:1101426
KEYWORDS EST.
SOURCE human.

[illegible]

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@edlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (http://www.tigr.org/tdb/humgen/bac_end_search). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search.html
Seq primer: 77
Class: BAC ends.

FEATURES
Source
Location/Qualifiers
1..219
/organism="Homo sapiens"
/db_xref="GDB:7529676"
/db_xref="taxon:9606"
/clone="RPC1-11-78E13"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; PRC111 Human Male BAC library"

BASE COUNT
69 a 45 c 46 g 59 t

ORIGIN

Query Match 1.9%; Score 58; DB 13; Length 219;
Best Local Similarity 100.0%; Pred. No. 8,1e-18;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2237 tgtacaacacctgaaacctgagtattcatcagaactcaagaatcaagcgtccgagt 2234
|||||
Db 104 TGTACAAACCTGAAACCTTGATGATTCATGACCTACAGAAATCAACGCGTCCGGGT 161

RESULT 15
B58691

LOCUS B58691 635 bp DNA GSS 20-JUN-1998
DEFINITION CIT-HSP-2013L21.TF CIT-HSP Homo sapiens genomic clone 2013L21, DNA sequence.
ACCESSION B58691
VERSION B58691.1 GI:2613409
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 635)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CIT-HSP-2013L21.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (http://www.tigr.org/tdb/humgen/bac_end_search). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.
Location/Qualifiers
1..635
/organism="Homo sapiens"
/db_xref="GDB:7042463"
/db_xref="taxon:9606"

FEATURES
Source

/clone="2013L21"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT 217 a 122 c 90 g 206 t
ORIGIN

Query Match 1.88; Score 55; DB 13; Length 635;
Best Local Similarity 100.0%; Pred. No. 2.7e-16;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2765 caaacctggcttgaagaactgcagacatcacagatcacagattgaattaggt 2819
|||||
Db 1 CAAAGCTGGGTTGAAAACCTGGAGACTCAGCATACAGACTTTAGATTAGGT 55

RESULT 16
LOCUS AI222422 446 bp mRNA EST 30-NOV-1998
DEFINITION qh04f06.x1 Soares_NFL_T-GBC_S1 Homo sapiens cDNA clone
IMAGE:1843715 3', mRNA sequence.
ACCESSION AI222422
VERSION AI222422.1 GI:3804625
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 446)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer genome anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 546 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 437.
Location/Qualifiers
1..446

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1843715"
/clone_1lb="Soares_NFL_T-GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH19W, testis NHT, and B-cell
NCI-CCAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1.M.A.G.E. clones 257480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Benito
Soares and M. Fatima Bonaldo.

BASE COUNT 128 a 95 c 86 g 137 t
ORIGIN

Query Match 1.0%; Score 32; DB 10; Length 446;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 298 gctctttcatcacatcagaagagactgt 329
|||||
Db 415 GTCCTTTTCATCAGCATCAGACGACTGTG 446

RESULT 17
LOCUS BF903662 251 bp mRNA EST 18-JAN-2001
DEFINITION IL2-MT0180-181200-276-F03 MT0180 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF903662
VERSION BF903662.1 GI:12295121
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 251)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL2&t=2-IL2-MT0180-
181200-276-F03&t=3-2000-12-18&t=1)
Seq primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 96.
Location/Qualifiers
1..251

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="MT0180"
/dev_stage="Adult"

/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
, 716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 72 a 62 c 56 g 61 t
ORIGIN

Query Match 1.0%; Score 30; DB 11; Length 251;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1092 gattccactctcacacacaacagctg 1121
|||||
Db 52 GAGTCCACCTCTCACACACAAACACGCTG 81

RESULT 18
LOCUS A0889169/c 404 bp DNA GSS 10-NOV-1999
DEFINITION HS-2161.B1.A01.T7C C17 Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2161 Col-1 Row=B, DNA sequence.
ACCESSION A0889169
VERSION A0889169.1 GI:6345359
KEYWORDS GSS.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 404)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (Info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 2161 Row: B Column: 1
Seq primer: T7
Class: BAC ends
High quality sequence stop: 404.
Location/Qualifiers
1..404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2161 Col=1 Row=B"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 119 a 93 c 70 g 122 t
ORIGIN

Query Match 0.9%; Score 28; DB 13; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 188 tagagggatcattcagatgattgaaa 215
|||||
Db 200 TAGAGGATCATTCACATGATTTGAAA 173

RESULT 19
LOCUS AO497022/c 443 bp DNA GSS 28-APR-1999
DEFINITION HS_5197_B1_C11.SP6E RPCR-11 Human Male BAC Library Homo sapiens genomic clone Plate-773 Col=21 Row=F, DNA sequence.
ACCESSION AO497022
VERSION AO497022.1 GI:4697145
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 443)
REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

FEATURES
SOURCE
Location/Qualifiers
1..674
/organism="Homo sapiens"

FEATURES
SOURCE
Location/Qualifiers
1..443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=773 Col=21 Row=F"
/clone_lib="RPCR-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 148 a 90 c 75 g 129 t 1 others
ORIGIN

Query Match 0.8%; Score 25; DB 13; Length 443;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 ggtgggaagcttcacgcagaaacaa 36
|||||
Db 180 GGTGGGAAGCTTCACGCAGAACAA 156

RESULT 20
LOCUS AO541295/c 674 bp DNA GSS 19-MAY-1999
DEFINITION RPCR-11-343A21.TV RPCR-11 Homo sapiens genomic clone RPCR-11-343A21, DNA sequence.
ACCESSION AO541295
KEYWORDS AO541295.1 GI:4871751
SOURCE GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 674)
REFERENCE Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCR-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other-GSS: RPCR-11-343A21.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCR-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet. cs (Info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.


```

/db_xref="GDB:7631348"
/db_xref="taxon:9606"
/clone="RPCL-11-343A21"
/clone_1lb="RPCL-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT      218 a      138 c      133 g      185 t
ORIGIN

```

```

Query Match      0.8%; Score 25; DB 13; Length 674;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

12 ggtgggaagcttcacagacaa 36
|||||
Db      170 GGTGGAGCTTCATCCAGACAA 146

RESULT 21
LOCUS      BF829853      606 bp      mRNA      EST      13-JAN-2001
DEFINITION MR3-HN0052-261200-002-cl2 HN0052 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF829853
VERSION     BF829853.1 GI:12175855
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

1 (bases 1 to 606)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=MR3&L2=MR3-HN0052-
261200-002-cl2&L3=2000-12-26&L4=1)
Seq primer: puc 18 forward
High quality sequence start: 24
High quality sequence stop: 538.
Location/Qualifiers

```

```

FEATURES
source
1. 606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="HN0052"
/dev_stage="Adult"
/notes="Organ: head_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent Application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

BASE COUNT      119 a      177 c      195 g      113 t      2 others
ORIGIN

```

```

Query Match      0.7%; Score 23; DB 11; Length 606;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

1922 agaacagcgtgcacatcg 1944
|||||
Db      100 AGACACAGGTGCATCCACATGG 122

RESULT 22
LOCUS      AV268403      256 bp      mRNA      EST      05-NOV-1999
DEFINITION AV268403 RIKEN full-length enriched, adult male testis (DH10B) Mus
musculus cDNA clone 4930534P12 3', mRNA sequence.
ACCESSION  AV268403
VERSION     AV268403.1 GI:6256440
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

```

1 (bases 1 to 256)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateo,M., Tomimaga,N., Tsunoda,Y.,
Wataniki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
url:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Wataniki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers

```

```

FEATURES
source
1. 256
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930534P12"
/clone_1lb="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Site_1: SalI; Site_2: BamHI; cDNA library was

```

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGGAGAGAGAGATCCACAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using Trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',
GAGGAGAGATTCGAGTATTAATATCCCGCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmid pUC19(+) after bulk excision from Lambda FLIC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 80 a 36 c 31 g 109 t

ORIGIN

Query Match 0.7%; Score 22; DB 10; Length 256;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2590 ttcttgattatcagaataa 2611
|||||
45 TTCTTGATTATCAGAAATTA 66

RESULT 23
LOCUS A2740007 437 bp DNA GSS 25-JAN-2001
DEFINITION RPCI-24-103A13.TV RPCI-24 Mus musculus genomic clone RPCI-24-103A13
, DNA sequence.
ACCESSION A2740007
VERSION A2740007.1 GI:12510653
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 437)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatman,S., Akintet,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorjais,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-103A13.TJB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cno.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 103 row: A column: 13
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. 437
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-103A13"
/clone_1id="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pFARBAC1; Site_1: BamHI; Site_2: BamHI.
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pFARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 194 a 81 c 63 g 99 t

ORIGIN

Query Match 0.7%; Score 22; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1151 gaaacaacacacacataa 1172
|||||
Db 204 GAAACAACACACAAACATMAA 225

RESULT 24
LOCUS A2086396 163 bp DNA GSS 08-MAY-2000
DEFINITION RPCI-23-26D1.TV RPCI-23 Mus musculus genomic clone RPCI-23-26D1,
DNA sequence.
ACCESSION A2086396
VERSION A2086396.1 GI:7728130
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 163)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatman,S., Akintet,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-26D1.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 26 row: D column: 1
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. 163
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-26D1"
/clone_1id="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. The selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 57 a 32 c 21 g 53 t

ORIGIN

Query Match 0.7%; Score 21; DB 13; Length 163;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 aaaaacaacacacataaa 1172
 Db 140 AAAACACACACACACATAA 160

RESULT 25
 AA784099 343 bp mRNA EST 29-JUL-1998
 LOCUS AA784099
 DEFINITION d1908a1.f1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans cDNA clone d1908a1 3', mRNA sequence.

ACCESSION AA784099
 VERSION AA784099.1 GI:2844267
 KEYWORDS EST.
 SOURCE Emericella nidulans.
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Emericella; Trichocomaceae; Emericella.

REFERENCE 1 (bases 1 to 343)
 Kupfer,D., Gray,D., Hausner,J., Lai,H., Martin,W., Aramayo,R., Prade,R., and Roe,B.
 An Aspergillus nidulans EST Database
 Unpublished (1998)
 Other ESTs: d1908a1.r1
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center
 Seq primer: M13-20
 High quality sequence stop: 301.
 Location/Qualifiers
 1..343
 /organism="Emericella nidulans"
 /strain="FGSC A26"
 /db_xref="taxon:162425"
 /clone_lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library"
 /tissue_type="vegetative mycelia, asexual structures"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 85 a 94 c 75 g 89 t

ORIGIN

Query Match 0.7%; Score 21; DB 10; Length 343;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2915 atttggaatcctaagaact 2935
 Db 19 ATTGGAATCTTAAGCAATT 39

RESULT 26
 AA821339 343 bp mRNA EST 17-FEB-1998
 LOCUS AA821339
 DEFINITION vs68a01.r1 Strataegene mouse skin (#937313) Mus musculus cDNA clone IMAGE:1151400 5', mRNA sequence.

ACCESSION AA821339
 VERSION AA821339.1 GI:2891207
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 343)
 AUTHORS Maria M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheißenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:624608
 Seg primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 336.
 Location/Qualifiers
 1..343
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone_lib="IMAG:1151400"
 /clone_lib="Strataegene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dr. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGACGACGACG 3' ~3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTT 3'"

BASE COUNT 84 a 103 c 95 g 61 t

ORIGIN

Query Match 0.7%; Score 21; DB 10; Length 343;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 787 atgaatcaagccccaact 807
 Db 319 ATGAATTCAGCCCGACACT 339

RESULT 27
 A1645009 371 bp mRNA EST 29-APR-1999
 LOCUS A1645009
 DEFINITION vs68a01.y1 Strataegene mouse skin (#937313) Mus musculus cDNA clone IMAGE:1151400 5', mRNA sequence.

ACCESSION A1645009
 VERSION A1645009.1 GI:4723484
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 371)
 AUTHORS Maria M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Maria M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mousest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MG:624608

This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)

Seq primer: -40RP from Gibco

High quality sequence stop: 355.

FEATURES

SOURCE

Location/Qualifiers

1..371

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:1151400"

/clone_lib="Stratagene mouse skin (#937313)"

/sex="females"

/tissue_type="whole skin"

/dev_stage="11 weeks old"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: skin; Vector: pBluescript SK-; Site:1: EcoRI

; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo

DT. Whole skin from 11 week old C57BL/6 female mice.

Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'

adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'

BASE COUNT
ORIGIN

92 a 109 c 101 g 69 t

Query Match 0.7%; Score 21; DB 10; Length 371;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 787 atgaattcaagccaggaact 807
|||||
Db 321 ATGAATTCAGCCCGAGACT 341

RESULT 28

AZ223093/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other GSS: RPCI-23-94H20.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaoc@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (Info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 94 row: H column: 20
Seq primer: 17
Class: BAC ends.

FEATURES
SOURCE

Location/Qualifiers

1..432

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-94H20"

/clone_lib="RPCI-23"

/sex="female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site:1:

EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT
ORIGIN

75 a 123 c 122 g 112 t

Query Match 0.7%; Score 21; DB 13; Length 432;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 787 atgaattcaagccaggaact 807
|||||
Db 115 ATGAATTCAGCCCGAGACT 95

RESULT 29

AO555443/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

or from Resea ch Genetics (Info@resgen.com). BAC end Web Server:

http://www.htsc.washington.edu

Plate: 800 row: C column: 10

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 437.

Location/Qualifiers

1..437

/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT      168 a      60 c      81 g      128 t
ORIGIN

Query Match      0.7%; Score 21; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1705 gtggcatccattatatacaag 1725
|||||
Db 387 gtggcatccattatatacaag 367

RESULT 30
CNS00SV2 462 bp DNA GSS 28-JUN-1999
CNS00SV2/c Arabidopsis thaliana genome survey sequence SP6 end of BAC T2N9 of
LOCUS TAMU library from strain Columbia of Arabidopsis thaliana, genomic
DEFINITION survey sequence.
ACCESSION AL088892.1 GI:5290032
VERSION AL088892.1
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 462)
REFERENCE Samson,D., Saurin,W., Artiguenave,F., Brotier,P., Wincker,P.,
AUTHORS Salanoubat,M., Choins,N.,
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 462)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
source
1.462
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_1lb="TAMU"
/clone="T2N9"
/notes="end : SP6"
BASE COUNT      129 a      87 c      70 g      176 t
ORIGIN

Query Match      0.7%; Score 21; DB 13; Length 462;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 626 caaatgctctctctctccg 646
|||||
Db 239 CAAATGCTCTCTCTCTCCG 219

RESULT 31
A2360053 602 bp DNA GSS 02-OCT-2000
LOCUS 1M0103H11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0103H11 F, DNA sequence.
ACCESSION A2360053

```

```

VERSION A2360053.1 GI:10473753
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 602)
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
AUTHORS Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0103 row: H column: 11
Seq primer: CGTTGTAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 602.
FEATURES
source
1.602
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC1M0103H11"
/clone_1lb="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/notes="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1147321149b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      139 a      153 c      139 g      171 t
ORIGIN

Query Match      0.7%; Score 21; DB 13; Length 602;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1745 cctgagccagaatttgaagc 1765
|||||
Db 21 CCGAGCCAGCAAGAAATGCAAC 1

RESULT 32
BG571228 1147 bp mRNA EST 10-APR-2001
LOCUS BG571228
DEFINITION 602592168F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4719424 5',
mRNA sequence.

```

ACCESSION BG571228
 VERSION BG571228.1 GI:13578881
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1147)
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM1572 row: h column: 17
 High quality sequence stop: 240.
 Location/Qualifiers
 1..1147
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4719424"
 /clone_id="NIH-MGC-79"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
 Site_1: Sfil (ggcgccctggcc); Site_2: Sfil (ggcgcattggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCCGCGCGCCGACATG-dt(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
 library."
 BASE COUNT 477 a 135 c 499 g 36 t
 ORIGIN
 Query Match 0.7%; Score 21; DB 11; Length 1147;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1148 acagaaaaaacacacaa 1168
 Db 298 ACAGAAAAACAAACACAAACA 318
 RESULT 33
 LOCUS B1142462 64 bp mRNA EST 05-JUL-2001
 DEFINITION SMOV3MCAM62G10SK Onchocerca volvulus molting L3 larva cDNA
 (SL96MLW-Ovml3) Onchocerca volvulus cDNA SMOV3MCAM62G10 5',
 mRNA sequence.
 ACCESSION B1142462
 VERSION B1142462.1 GI:14624172
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
 REFERENCE 1 (bases 1 to 64)
 AUTHORS Williams, S.A., Lizotte-Waniewski, M., Ianey, S. and Lustigman, S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pBluescript-SK.
 Location/Qualifiers
 1..64
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone_image="SMOV3MCAM62G10"
 /clone_id="Onchocerca volvulus molting L3 larva cDNA
 (SL96MLW-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS
 in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
 culture. L3 of O. volvulus molt to fourth-stage larvae by
 day 5 in culture. mRNA was isolated from approximately
 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
 in culture, and converted to double-stranded cDNA using
 reverse transcriptase and oligo(dt) followed by RNase H
 and DNA pol I. The library was constructed in the lambda
 Uni-Zap XR vector and has 1 x 10⁶ independent
 recombinants and the average insert size is ~1200 bp. The
 library was constructed by Sara Lustigman and Michelle
 Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
 The library is available from Dr. Sara Lustigman (email:
slustigmen@bc.org)."
 BASE COUNT 17 a 7 c 20 g 20 t
 ORIGIN
 Query Match 0.6%; Score 20; DB 11; Length 64;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2364 gaagaagatgcataaact 2383
 Db 42 GAAGAAGATGCTATTAACCT 61
 RESULT 34
 LOCUS BE638399 73 bp mRNA EST 28-AUG-2000
 DEFINITION SMOVL2CASI5B03SK Onchocerca volvulus L2 larvae cDNA (SAW98MLW-OvL2)
 Onchocerca volvulus cDNA SMOVL2CASI5B03 5', mRNA sequence.
 ACCESSION BE638399
 VERSION BE638399.1 GI:9937018
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
 REFERENCE 1 (bases 1 to 73)
 AUTHORS Williams, S.A.
 TITLE Genes expressed in L2 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1999)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pBluescript SK.
 Location/Qualifiers
 1..73

/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOVL2CASI5B03"
/clone_lib="Onchocerca volvulus L2 larvae cDNA
(SAM96LM-OVL2)"
/dev_stage="L2"
/lab_host="X11-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. mRNA was
prepared from approximately 9,000 L2s isolated from
infected mosquitoes from Kumba, Cameroon and converted to
double-stranded cDNA using reverse transcriptase and
oligo(dT) followed by RNase H and DNA pol I. The library
has 7.3 x 10E4 independent recombinants and the average
insert size is approximately 1kb. The library was
constructed by Michelle Lizotte-Waniewski. The library is
available from Dr. S.A. Williams, email: genomesmith.edu."

BASE COUNT 22 a 11 c 19 g 21 t

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2364 gaagaagatcgtctaaact 2383
|||||

Db 27 GAAGAGATCTCTATAACT 46

RESULT 35
LOCUS A0069062 205 bp DNA GSS 04-AUG-1998
DEFINITION HS-2225-B2_H03_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2255 Col=6 Row=P, DNA sequence.
ACCESSION A0069062
VERSION A0069062.1 GI:3384261
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 205)
Mahatras G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahatras G.G., Wallace J.C., Hood L.
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2255 Row: P Column: 6
Class: BAC ends
High quality sequence stop: 205.
Location/Qualifiers
1. 205
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2255 Col=6 Row=P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 54 a 41 c 52 g 57 t 1 others

ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1147 tacagaacaacacacaa 1166
|||||

Db 134 TACAGAAACAAACACAA 115

RESULT 36
LOCUS B1315256 215 bp mRNA EST 20-JUL-2001
DEFINITION dah9e06.y1 NICHD XGC Emb4 Xenopus laevis cDNA clone IMAGE:4957906
5' mRNA sequence.
ACCESSION B1315256
VERSION B1315256.1 GI:14989583
KEYWORDS EST.
SOURCE African clawed frog.
Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 215)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 197.
Location/Qualifiers
1. 215
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4957906"
/clone_lib="NICHD XGC Emb4"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/note="Organ: whole embryo; Vector: PCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
Library.)"

BASE COUNT 72 a 32 c 37 g 74 t

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2825 ttctggaagaacctctga 2844
|||||

Db 196 TTTTGAAGAACACCTCTGA 215

RESULT 37
LOCUS AA493806 216 bp mRNA EST 19-AUG-1997
DEFINITION nh02f04.s1 NCI-CGAP_Thy1 Homo sapiens cDNA clone IMAGE:943135, mRNA
sequence.
ACCESSION AA493806
VERSION AA493806.1 GI:2223647
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 216)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 294 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES
source Location/Qualifiers
1..216
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="943135"
/clone_lib="NCI-CGAP_Thy1"
/tissue_type="thyroid"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from invasive thyroid
tumor, CDNA made by oligo-dT priming. Non-directionally
cloned. Size selected on agarose gel, average insert size
600 bp. Reference: Kitzman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 57 a 65 c 37 g 57 t
ORIGIN
Query Match 0.68; Score 20; DB 10; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2225 gcaatcacatctgaacaa 2244
|||||
Db 92 GCACATCACATCTGAACAA 111
RESULT 38
AI366637 225 bp mRNA EST 08-JAN-1999
LOCUS SMOV3MCAM11C04SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLW-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCAM11C04 5',
mRNA sequence.
ACCESSION AI366637
VERSION AI366637.1 GI:4136382
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 225)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
FEATURES
Location/Qualifiers

source
1..225
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM11C04"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLW-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-Zap XR; Site 1: Eco RI; Site 2:
Xho I; Filarial nematode parasite of humans, third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3) 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nybc.org)."
BASE COUNT 65 a 39 c 57 g 64 t
ORIGIN
Query Match 0.68; Score 20; DB 10; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2364 gaagagaatgctataaact 2383
|||||
Db 30 GAAGAGATGCTATTAACCT 49
RESULT 39
AI603832 248 bp mRNA EST 21-APR-1999
LOCUS SMOV3MCAM27A08SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLW-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCAM27A08 5',
mRNA sequence.
ACCESSION AI603832
VERSION AI603832.1 GI:4612981
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 248)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
FEATURES
source Location/Qualifiers
1..248
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM27A08"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLW-Ovml3)"


```

/dev-stage="molting L3"
/lab-host="XLI-Blue MRF"
/notes="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroun (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 106 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Manlewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigmen@bc.org)."

```

BASE COUNT 71 a 47 c 58 g 69 t 3 others

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 248;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2364 gaagaagatctataaact 2383
 ||||||||||||||||
 Db 14 GAAGAAGATCTATAAAACT 33

RESULT 40
 LOCUS AZ891870 299 bp DNA GSS 05-MAR-2001
 DEFINITION RPCI-24-180D14.TJ RPCI-24 Mus musculus genomic clone RPCI-24-180D14
 , DNA sequence.
 ACCESSION AZ891870
 VERSION AZ891870.1 GI:13210815
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 299)
 Zhao, S., Niernan, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shwartsdeyn, A., Georegeorgis, E.,
 Russell, D., de Jong, P., and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other GSSs: RPCI-24-180D14.TJ
 COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: sshaet@igr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pjejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 180 row: D column: 14
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..299
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-24-180D14"
 /clone_lib="RPCI-24"

FEATURES
 source

1..299
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-24-180D14"
 /clone_lib="RPCI-24"

```

/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI;  

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  

library was cloned in the pPARBAC1 cloning vector at the  

BamHI sites using MboI partially digested male C57BL/6J  

DNA."

```

BASE COUNT 103 a 57 c 85 g 54 t

ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 299;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 989 tgaagctgtgtctccaa 1008
 ||||||||||||||||
 Db 202 TGAAGCTGTGTCTCCAA 221

RESULT 41
 LOCUS AW357412 332 bp mRNA EST 25-APR-2001
 DEFINITION 40862 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AW357412
 VERSION AW357412.1 GI:6861490
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 332)
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
 Casas, E., Wray, J.E., White, D., Cho, J., Fahrnenkrug, S.C., Bennett,
 G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
 Perta, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
 Keeler, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -mnscore 20
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCAGTCAGCAGCAG
 Plate: 20 row: J column: 17
 Seq primer: ATTGAGTGACACTATAG.
 Location/Qualifiers
 1..332
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10b"
 /note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."

BASE COUNT 59 a 105 c 88 g 80 t

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 733 tggccatgctgctgaagctg 752
 |||
 Db 77 TGGCCATGCTGCTGAAGCTG 96

RESULT 42

A0952434/c

LOCUS 336 bp DNA 27-JAN-2000
 DEFINITION Shared DNA-48N4-TR Sheared DNA Trypanosoma brucei genomic clone

ACCESSION A0952434
 VERSION A0952434.1 GI:6775699

KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei

TITLE Trypanosoma. Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 JOURNAL Trypanosoma.

REFERENCE 1 (bases 1 to 336)
 AUTHORS El-Sayed, N., Zhao, H., Gill, S., Sub, E., Malek, J., Fujii, C.,
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
 Fraser, C. and Adams, M.

COMMENT Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library
 Unpublished (1999)
 Other_GSSs: Sheared DNA-48N4-TR
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@ligr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through ATCC. Sheared DNA end sequences search page:
 http://www.ligr.org/cdb/mdb/cdb/.
 Seq primer: M13-Reverse
 Class: shotgun.

FEATURES
 Location/Qualifiers

1..336
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-48N4"
 /clone_lib="Sheared DNA"

/note="Vector: pUC18; Site_1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Yaudin and B. Bartell, Oxford University
 Press, 1999)."

BASE COUNT 72 a 62 c 68 g 134 t
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 336;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1153 aaacaacacacataaa 1172
 |||
 Db 171 AAAACAACACAAACATMAA 152

RESULT 43
 AA670222/c

LOCUS AA670222 342 bp mRNA EST 20-NOV-1997
 DEFINITION ad19hl.s1 Soares_NBHF Homo sapiens cDNA clone IMAGE:878757 3',
 mRNA sequence.
 ACCESSION AA670222
 VERSION AA670222.1 GI:2631721
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 342)
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Matra, M., Mattin
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
 White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 CONTACT: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 334.

FEATURES
 Location/Qualifiers

1..342
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:878757"
 /clone_lib="Soares_NBHF"
 /dev_stage="15 wk post natal"
 /lab_host="DH10B"

/note="Organ: whole brain; Vector: p773D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 15' AACTGGAAGATTCGCGCGCAATATTTTATTTTATTTT 3'.
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 72 a 122 c 81 g 67 t
 ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 648 ctccacagggcccgaggtg 667
 |||
 Db 137 CTCACAGGCGCCAGGtG 118

RESULT 44

A2991565

LOCUS A2991565 380 bp DNA 27-APR-2001
 DEFINITION 2M0275B24R Mouse 10kb plasmid U062M library Mus musculus genomic
 clone U062M0275B24 R, DNA sequence.

ACCESSION A2991565
 VERSION A2991565.1 GI:13862792

KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 380)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

TITLE and Wright, D., Weiss, R.
JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb
COMMENT plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0275 row: B column: 24
 Seq primer: CACACGGAACACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 380.

FEATURES

source
 1. 380
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U06C2M0275B24"
 /clone_lib="Mouse 10kb plasmid U06C2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

140 a 71 c 97 g 72 t
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 380;
Best local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1153 aaacaacaacaacataaa 1172
 ||||||||||||||||||||
Db 35 AAAACAACACCAACATTA 54

RESULT 45
AA294641 393 bp mRNA EST 12-NOV-1997
LOCUS SM0V3MCA885SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone Smml3C0885 5', mRNA sequence.
ACCESSION AA294641
VERSION AA294641.1 GI:2099635
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 393)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.

TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genomes@smith.edu
 Seq primer: pBluescript SK.

FEATURES

source
 1. 393
 Location/Qualifiers
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="Smml3C0885"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCIC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@ncic.org)."

BASE COUNT

104 a 79 c 96 g 111 t 3 others
 ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 393;
Best local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgctataaact 2383
 ||||||||||||||||||||
Db 42 GAAGAAGATGCTATAAACT 61

RESULT 46
AM838955 393 bp mRNA EST 18-MAY-2000
LOCUS CM2-LT0061-180200-094-h05 LT0061 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM838955
ACCESSION AM838955
VERSION AM838955.1 GI:7932929
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 393)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-cm2-LR0061-180
200-094-H054t3-2000-02-18&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 393.

FEATURES

source

```
1..393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LR0061"
/dev_stage="Adult"
/note="Organ: leiomyos; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      90 a      99 c      115 g      89 t
ORIGIN
```

Query Match 0.6%; Score 20; DB 10; Length 393;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 tactattgactctctga 1834

|||||

Db 68 TACTTATTGACTCTTCTTGA 49

RESULT 47

AA294540

LOCUS SMOV3MCA61SK Onchocerca volvulus molting L3 larva cDNA 12-NOV-1997

DEFINITION (SI96MLM-OvMLJ) Onchocerca volvulus cDNA SMOV3C0761 5', mRNA

SEQUENCE.

ACCESSION AA294540

VERSION AA294540.1 GI:2099534

KEYWORDS EST.

SOURCE Onchocerca volvulus.

ORGANISM

Onchocerca volvulus

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;

Onchocercidae; Onchocerca.

1 (bases 1 to 415)

Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.

Genes expressed in molting L3 larvae of Onchocerca volvulus

Unpublished (1997)

Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: PBluescript SK.

Location/Qualifiers

1..415

/organism="Onchocerca volvulus"

/strain="Kumba, Camerouns"

/db_xref="taxon:6283"

/clone="SMML3C0761"

/clone_lib="Onchocerca volvulus molting L3 larva cDNA

(SI96MLM-OvMLJ)"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in

Cameron (Forest strain). The L3 were cultured in 20% FCS

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in

culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately

6000 molting larvae (MLJ). 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using

reverse transcriptase and oligo(dT), followed by RNase H

and DNA pol I. The library was constructed in the Lambda

Uni-Zap XR vector and has 1 x 10⁶ independent

recombinants and the average insert size is ~1200 bp. The

library was constructed by Sara Lustigman and Michelle

Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email:

slustigman@bc.org)."

BASE COUNT 110 a 89 c 103 g 102 t 11 others

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 415;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgctataaact 2383

|||||

Db 17 GCACGACGATGCTATTAACCT 36

RESULT 48

T81898/c

LOCUS T81898 415 bp mRNA EST 15-MAR-1995

DEFINITION yd29a11.s1 Soares fetal liver spleen INFILS Homo sapiens cDNA clone

IMAGE:109628 3', mRNA sequence.

ACCESSION T81898

VERSION T81898.1 GI:704905

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 415)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,

M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston,

R., Williamson, A., Woldmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Other ESTs: yd29a11.r1

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 714

High quality sequence stops: 314 Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 714 Std Error: 0.00

Seq primer: T3

High quality sequence stop: 314.

Location/Qualifiers

1..415

/organism="Homo sapiens"

/db_xref="GDB:465245"

/db_xref="taxon:9606"

/clone="IMAGE:109628"

/clone_lib="Soares fetal liver spleen INFILS"

BASE COUNT	112 a	91 c	107 g	123 t	6 others
ORIGIN					
Query Match	0.6%; Score 20; DB 10; Length 439;				
Best Local Similarity	100.0%; Pred. No. 1.3e+02;				
Matches 20; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
OY	2364 gaagaagatgctataaac	2383			
Db	42 GAAGAAGATGCTATAAAACT	61			
IIIIIIIIIIIIIIIIIIII					
RESULT 50					
LOCUS	AO209355				
DEFINITION	AO209355 483 bp DNA GSS 18-SEP-1998				
ACCESSION	HS_3240.A2.B12.T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3240 Col=24 Row=C, DNA sequence.				
VERSION	AO209355.1 GI:3622090				
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 483)				
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T', Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to maping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)				
MEDLINE	99380589				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3240 row: C column: 24 Class: BAC ends High quality sequence stop: 483. Location/Qualifiers 1..483 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="plate=3240 Col=24 Row=C" /clone_id="CIT Approved Human Genomic Sperm Library D" /sex="male" /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"				
BASE COUNT	137 a	135 c	102 g	108 t	1 others
ORIGIN					
Query Match	0.6%; Score 20; DB 13; Length 483;				
Best Local Similarity	100.0%; Pred. No. 1.3e+02;				
Matches 20; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
OY	1845 aatgtgtcaagtgtctctgga	1864			
Db	324 AATTGTCAAGTGTCTCTGGA	343			
IIIIIIIIIIIIIIIIIIII					
RESULT 51					
PF601894/C					

LOCUS BF601894 491 bp mRNA EST 25-APR-2001
 DEFINITION 266990 MARC 3BOV Bos taurus CDNA 5', mRNA sequence.
 ACCESSION BF601894
 VERSION BF601894.1 GI:11699117
 KEYWORDS EST.
 SOURCE
 ORGANISM Bos taurus
 Cow.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Pettea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine CDNA libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
 PCR PRIMERS
 FORWARD: AGGAACACGCTATGACCAT
 BACKWARD: GTTTCACGTCACGACG
 Plate: 42 row: N column: 11
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1. 491
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."
 BASE COUNT 120 a 104 c 105 g 162 t
 ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 491;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2787 agactcacagatcacagat 2806
 |||||||
 DB 284 AGACTCACAGATCACAGAT 265

RESULT 52
 LOCUS A1643174 498 bp mRNA EST 29-APR-1999
 DEFINITION V446909.Y1 Soares_mammary_gland_NbMKG Mus musculus cDNA clone
 IMAGE:864448 5', mRNA sequence.
 ACCESSION A1643174
 VERSION A1643174.1 GI:4721649
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 498)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
 COMMENT
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:508536
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the correct orientation)
 Seq primer: -40RP from Gbpco
 High quality sequence stop: 484.
 Location/Qualifiers
 1. 498
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:864448"
 /clone_lib="Soares_mammary_gland_NbMKG"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGCGCGCGAGGCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 138 a 130 c 108 g 121 t 1 others
 ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 498;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1327 cagctcaagggttcagcca 1346
 |||||||
 DB 498 CAGCTCAAGGTTCAGCCA 479

RESULT 53
 LOCUS BF602580 505 bp mRNA EST 25-APR-2001
 DEFINITION 267845 MARC 3BOV Bos taurus CDNA 5', mRNA sequence.
 ACCESSION BF602580
 VERSION BF602580.1 GI:11699804
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Cow.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Pettea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine CDNA libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366

Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

FEATURES

Source

1. 505
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
Location/Qualifiers

BASE COUNT

96 a 135 c 132 g 122 t

ORIGIN

Query Match

0.6%; Score 20; DB 11; Length 505;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

Qy 733 tggccatctgtctgaagctg 752
|||||

Db 174 TGCCATGCTGCTGACACTG 193

RESULT 54
BG382487 519 bp mRNA EST 12-MAR-2001
LOCUS
DEFINITION 298398 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BG382487
VERSION BG382487.1 GI:13306959
KEYWORDS EST.
SOURCE
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 519)
Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

FEATURES

Source

1. 519
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos "
BASE COUNT 150 a 103 c 104 g 162 t
ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 519;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

Qy 1823 tgactcttgaacattgc 1842
|||||

Db 312 TGACTCTTGAACATTGTC 293

RESULT 55

LOCUS

A1670486 521 bp mRNA EST 17-MAY-1999
DEFINITION SMOV3MCM26G12SK Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCM26G12 5',
mRNA sequence.

ACCESSION

A1670486
A1670486
A1670486.1 GI:4836992

VERSION

KEYWORDS

SOURCE

ORGANISM

EST.

Onchocerca volvulus.

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;
Onchocercidae; Onchocerca.
1 (bases 1 to 521)
Williams,S.A., Lizotte-Maniowski,M., Laney,S. and Lustigman,S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

1. 521

/organism="Onchocerca volvulus"

/strain="Kumba, Cameroons"

/db_xref="taxon:6282"

/clone_lib="SMOV3MCM26G12"

/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3)"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRP"

/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10⁶ independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Maniowski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigman@bc.org)."

BASE COUNT

139 a 105 c 125 g 147 t 5 others

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2364 gaagaagatgcataaact 2383
|||||

DB 42 GAAGAAGATGCTATAAACT 61

RESULT 56
FR0031506/c 521 bp DNA GSS 27-JUN-1998
LOCUS Fugu rubripes GSS sequence, clone 116B21aA4, genomic survey
DEFINITION Fugu rubripes GSS sequence, clone 116B21aA4, genomic survey
sequence.
ACCESSION AL027875.1 GI:3269989
VERSION AL027875.1 GI:3269989
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 521)
Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrandia, Y.,
Williams, G. and Brenner, S.
Direct Submission
Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridgeshire, CB10 1SB, UK. Email:
biohelp@hgm.mrc.ac.uk
Vector: pBluescript II KS
V-type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES
source Location/Qualifiers
1..521
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 116B21"
/clone="116B21aA4"

BASE COUNT 125 a 120 c 120 g 136 t 20 others

ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2620 aagatggaatgaagctctt 2639
|||||

DB 312 AAGATGAATGAGCTCTT 293

RESULT 57
BI394563 526 bp mRNA EST 06-AUG-2001
LOCUS pcp1n.pk001.110 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library Gallus gallus cDNA clone pcp1n.pk001.110 5' similar to
g117710042 ref1np.057930.11 IQ motif containing GTPase activating
protein 1; Cdc42-Rac1 effector protein [Mus musculus]
sp109URF110G1.MOUSE RAS GTPASE-ACTIVATING-LIKE PROTEIN IQGAPI
gblAA660344.1|AF240630.1 (AF240630) IQ motif containing GTPas, mRNA
sequence.
ACCESSION BI394563
VERSION BI394563.1 GI:15087845
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 526)
AUTHORS Porter, T.E. and Cogburn, L.A.
TITLE ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library USDA/IFARS Animal Genome Project
JOURNAL Unpublished (2001)
COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburnudel.edu, www.chickest.udel.edu.

FEATURES
source Location/Qualifiers
1..526
/organism="Gallus gallus"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="pcp1n.pk001.110"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5
w7,w9)."
/lab_host="E. Coli EMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"
BASE COUNT 124 a 162 c 151 g 87 t 2 others

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 526;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 740 gctgctgaagctgcgcaga 759
|||||

DB 72 GCTGCTGAAGCTGCGCAGA 91

RESULT 58
BG351255 535 bp mRNA EST 01-MAR-2001
LOCUS 08BD06 Mature tuber lambda ZAP Solanum tuberosum cDNA, mRNA
DEFINITION BG351255
ACCESSION BG351255
VERSION BG351255.1 GI:13179997
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 535)
Nielsen, K.L., Crookshanks, M., Emmersen, J. and Wellinder, K.G.
EST-sequencing of mature potato tuber (Var. Kurat)
Unpublished (2000)
Contact: Karen G. Wellinder
Institut for bioteknologi
Aalborg Universitet
Sohnsgaardsholmsvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808
Email: kgw@bio.auc.dk
Sequenced from the 5' end.
High quality sequence stop: 535
POLYA-No.

FEATURES
source Location/Qualifiers
1..535
/organism="Solanum tuberosum"
/cultivar="Field grown Kurat"
/db_xref="taxon:4113"
/clone_lib="Mature tuber lambda ZAP"

BASE COUNT 177 a 110 c 109 g 139 t
 ORIGIN /tissue_type="Tuber"
 /note="Vector: Lambda ZAP"

Query Match 0.6%; Score 20; DB 11; Length 535;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2395 tgaataactgaagaagatg 2414
 |||||||
 Db 160 TGAATAACTGAGAAGATG 179

LOCUS 59
 22190 A221490 541 bp DNA GSS 14-JUN-2000

DEFINITION Gm_UMB001_125_D12.R UMN Soybean BAC Library (PECSBAC4 EcORI)
 Glycine max genomic clone Gm_UMB001_125_D12, DNA sequence.

ACCESSION A221490
 VERSION A221490.1 GI:8517274
 KEYWORDS GSS.

SOURCE soybean.
 ORGANISM Glycine max

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

1 (bases 1 to 541)
 AUTHORs Marek, L.F., Paz, M., Darnielle, L., Hanson, N. and Shoemaker, R.C.
 TITLE BAC End sequences from a soybean genomic library (ISU)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Shoemaker Randy C
 Agronomy Department
 Iowa State University
 Ames, IA 50011-1010, USA
 Tel: 515 294 1205
 Fax: 515 294 2299
 Email: rcschoe@iastate.edu

FEATURES
 source This BAC identified by SSR Sat427. For more information, see
 Soybase at:
 http://genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase.
 Please see as an authority for the mapping/naming: Cregan P.B., T.
 Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya,
 T.T. Vantaoi, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An
 integrated genetic linkage map of the soybean genome. Crop Sci.
 39:(in press)
 Seg primer: BACR or M13R
 Class: BAC ends.

FEATURES
 source Location/Qualifiers

1..541
 /organism="Glycine max"
 /cultivar="Faribault"
 /db_xref="taxon:3847"
 /clone="Gm_UMB001_125_D12"
 /clone_11b="UMN Soybean BAC Library (PECSBAC4 EcORI)"
 /tissue_type="cotyledon leaves"
 /dev_stage="cotyledon stage"
 /note="Vector: PECSBAC4; The UMN BAC library (Danesh et al
 /Theor. Appl. Genet. 96:196, 1998) was constructed using
 the Eco RI site of PECSBAC4. The library consists of 72
 ,960 clones with an average insert size of 120 kb, equal
 to 7 haploid genome equivalents. Screening of the library
 is done by PCR amplification of DNA pools."
 BASE COUNT 177 a 88 c 90 g 165 t 21 others
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 541;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1024 gcttgagaatcatcatgaag 1043
 |||||||
 Db 424 GCTTGAGAACTCATCATGAG 443

RESULT 60
 A2298434 554 bp DNA GSS 27-JUL-2000
 LOCUS RPCI-23-109A14.TV RPCI-23 Mus musculus genomic clone RPCI-23-109A14
 DEFINITION , DNA sequence.

ACCESSION A2298434
 VERSION A2298434.1 GI:9540219
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 554)
 AUTHORs Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret,
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-23-109A14.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 109 row: A column: 14
 Seg primer: T7
 Class: BAC ends.

FEATURES
 source Location/Qualifiers

1..554
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-109A14"
 /clone_11b="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
 EcORI; Site 2: EcORI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcORI and EcORI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcORI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 179 a 80 c 76 g 219 t
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 554;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1868 caataactgaccttattg 1887
 |||||||
 Db 97 CATTAACGTGACTTTATG 116

RESULT 61
 A0270247 557 bp DNA GSS 03-NOV-1998
 LOCUS A0270247/c
 DEFINITION HS_2045_A1_G04_MR CIT Approved Human Genomic Sperm Library D Homo

ACCESSION AQ270247.1 GI:3822842
 VERSION AQ270247.1 GI:3822842
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 557)
 AUTHORS Maitiras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Maitiras GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2045 row: M column: 7
 Class: BAC ends
 High quality sequence stop: 557.
 Location/Qualifiers
 1..557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2045 Col=7 Row=M"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 137 a 140 c 108 g 169 t 3 others
 ORIGIN
 Query Match 0.6%; Score 20; DB 13; Length 557;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1940 catggaagagggccagagaa 1959
 ||||||||||||||||||
 Db 192 CATGGAAGAGGCCCGAGAA 173
 RESULT 62
 BF258009/c 571 bp mRNA EST 23-FEB-2001
 LOCUS HVSMET0014H19f Hordeum vulgare seedling root EST library HVCDNM0007
 DEFINITION (etiolated and unstressed) Hordeum vulgare cDNA clone
 HVSMET0014H19f, mRNA sequence.
 ACCESSION BF258009
 VERSION BF258009.2 GI:13119185
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 571)
 AUTHORS Wing,R., Close,T.J., Kleinjohs,A., Wise,R., Begum,D., Fritsch,D., Yu,X., Anderson,H., Dale,J., Henry,D., Kennel,S., Palmer,M., Rambo,T., Sasaki,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and Wood,T.
 TITLE Development of a genetically and physically anchored EST resource for barley genomics
 JOURNAL Unpublished (2000)
 COMMENT On Nov 16, 2000 this sequence version replaced gi:11187122.
 CONTACT: Wing RA

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: AATTAACCTCAGCTAAAGG
 High quality sequence stop: 558.
 Location/Qualifiers
 1..571
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMET0014H19f"
 /clone_lib="Hordeum vulgare seedling root EST library HVCDNM0007 (etiolated and unstressed)"
 /tissue_type="Seedling root"
 /lab_host="TJC121"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley/ To order a clone see http://www.genome.clemson.edu/orders"
 BASE COUNT 101 a 166 c 218 g 86 t
 ORIGIN
 Query Match 0.6%; Score 20; DB 11; Length 571;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 500 tggcctctgcagctcttc 519
 ||||||||||||||||||
 Db 191 TGGCCTCTCGAGGCTCTTC 172
 RESULT 63
 BG385206/c 577 bp mRNA EST 12-MAR-2001
 LOCUS BG385206 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION BG385206
 VERSION BG385206.1 GI:13309678
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 577)
 AUTHORS Fahrenkrug,S.C., Fieking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keeler,J.W.
 TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
 JOURNAL Unpublished (2000)
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smiththe@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
 PCR primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTTCCAGTACGACG
 Plate: 94 row: J column: 10
 Seg primer: ATTGAGTGACATATAG.
 Location/Qualifiers
 1..577
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 1Pig"
 /tissue_type="pooled"

/lab_host="DH10B"
 /note="Vector: PCMW SPORF6; Site_1: XbaI; Site_2: XhoI;
 library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."
 BASE COUNT 102 a 200 c 175 g 100 t
 ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 577;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 420 aactgaagaacacccctcac 439
 572 AACTGAAAGACACCTTCAC 553

RESULT 64
 LOCUS A0570315 577 bp DNA GSS 01-JUN-1999
 DEFINITION HS.5352.B1.E02.T7A RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=928 Col=3 Row=J, DNA sequence.
 ACCESSION A0570315
 VERSION A0570315.1 GI:4963535
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 577)
 Mahitras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahitras GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm)
 or from Research Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 928 row: J column: 3
 Seq primer: 17
 Class: BAC ends
 High quality sequence stop: 577.
 Location/Qualifiers

FEATURES
 source 1..577
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=928 Col=3 Row=J"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: PBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 PBACE3.6 vector at EcoRI sites"
 BASE COUNT 244 a 114 c 82 g 130 t 7 others
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 577;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2105 cagcgcacaataaagat 2124
 348 CAGCGTGCATAATAAGAGAT 367

RESULT 65
 LOCUS A2987277/c 580 bp DNA GSS 27-APR-2001
 DEFINITION 2M0269J15R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0269J15 R, DNA sequence.
 ACCESSION A2987277
 VERSION A2987277.1 GI:13858504
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 580)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 1000 Std Error: 0.00
 Plate: 0269 row: J column: 15
 Seq primer: CACACAGCAACAGCATGACCC
 Class: Plasmid ends
 High quality sequence stop: 580.
 Location/Qualifiers

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 1000 Std Error: 0.00
 Plate: 0269 row: J column: 15
 Seq primer: CACACAGCAACAGCATGACCC
 Class: Plasmid ends
 High quality sequence stop: 580.
 Location/Qualifiers

FEATURES
 source 1..580
 Location/Qualifiers

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0269J15"
 /clone_lib="Mouse 10kb plasmid UUGC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PM042ny; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PM042 (9114732114[9b]AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
 BASE COUNT 169 a 91 c 227 t
 ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 580;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1338 ttcaagccaagtaaatc 1357
 |||||
 Db 479 TTCAGCCCAAGTATAAT 460

RESULT 66 BG098273

LOCUS BG098273 605 bp mRNA EST 29-JAN-2001
 DEFINITION EST462792 sprouting eyes/shoots Solanum tuberosum cDNA clone
 C57C2015 5' sequence, mRNA sequence.

ACCESSION BG098273
 VERSION BG098273.1 GI:12586308

KEYWORDS SOURCE

ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 605)
 van der Hoeven,R.S., Bezzerides,J., Cho,J., Uterback,T., Hansen
 ,C.L., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker

REFERENCE AUTHORS

TITLE Generation of ESTs from potato sprouting eyes/shoots
 JOURNAL Unpublished (2001)
 COMMENT Contact: Cathy Ronning
 The Institute for Genomic Research
 For clone info: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdna@resgen.com.

FEATURES

Source

1..605

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="C57C2015"

/clone_lib="sprouting eyes/shoots"

/tissue_type="sprouting tubers"

/dev_stage="12-14 weeks post harvest"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:

XhoI: Various sizes of sprouting eyes (2 mm to 15 mm) were

taken from tubers. The tubers were incubated at 26C in

the dark for 2-3 weeks prior to sprouting. The eyes were

frozen in liquid nitrogen immediately upon removal from

tubers."

BASE COUNT 192 a 117 c 119 g 177 t

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 605;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2395 tgaataacctgaagaagt 2414

|||||

Db 156 TGAATAACCTGAGAGATG 175

RESULT 67 A2802117

LOCUS A2802117 620 bp DNA GSS 16-FEB-2001
 DEFINITION 2M0061H04F Mouse 10kb plasmid U06C1M library Mus musculus genomic
 clone U06C2M0061H04 F, DNA sequence.

ACCESSION A2802117
 VERSION A2802117.1 GI:12954536

KEYWORDS SOURCE

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 620)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

TITLE

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

CONTACT: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0061 row: H column: 04

Seq primer: CGTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 620.

FEATURES source

1..620

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="U06C2M0061H04"

/clone_lib="Mouse 10kb plasmid U06C1M library"

/sex="Male"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42n; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PMD42 (91473211419b/AP129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 167 a 141 c 175 g 137 t

ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 620;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2145 ctcaagttggtctcagcac 2164

|||||

Db 180 CTCAGTTGTCCTCAGCAC 199

RESULT 68 A1511575

LOCUS A1511575 627 bp mRNA EST 16-MAR-1999
 DEFINITION SMOVL3CAN30H03SK Onchocerca volvulus infective larva cDNA
 (SAM94WL-OV13) Onchocerca volvulus cDNA clone SMOVL3CAN30H03 5',

mRNA sequence.

ACCESSION A1511575
 VERSION A1511575.1 GI:4417429

KEYWORDS SOURCE

ORGANISM Onchocerca volvulus.

Onchocerca volvulus.

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;

Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 627)
AUTHORS Williams, S.A., Lu, W., Lizotte-Maniewski, M. and Laney, S.J.
TITLE Genes expressed in infective third stage larvae of *Onchocerca volvulus*
JOURNAL Unpublished (1995)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.
Location/Qualifiers
1..627
/organism="Onchocerca volvulus"
/strain="Sierra Leone"
/db_xref="taxon:6282"
/clone="SMOVL3CAN30H03"
/clone_1ib="Onchocerca volvulus infective larva cDNA (SAM94WL-OVL3)"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda UniZap XR; Site 1: EcoR I; Site 2: Xho I; Cutaneous filarial nematode parasite of humans. *Onchocerca volvulus* isolated from mosquitoes 10 days after infection and converted to double stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNase I. The library had 1.8 x 10E5 independent recombinants and average insert size was 900 base pairs. The library was constructed by Wenhong Lu. The library is available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 173 a 123 c 148 g 183 t
ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 627;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383
|||||
b 28 GAAGAAGATGCTATAAACT 47

RESULT 69
A2510047 633 bp DNA GSS 05-OCT-2000
LOCUS 1M0354H1AF Mouse 10kb plasmid UUGCIM library Mus musculus genomic
DEFINITION clone UUGCIM0354H14 F, DNA sequence.
ACCESSION A2510047
VERSION A2510047.1 GI:10691363
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 633)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellily, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

JOURNAL COMMENT

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0354 row: H column: 14
Seq primer: CGTTGTAACGACGGCCACT
Class: Plasmid ends
High quality sequence stop: 633.
Location/Qualifiers
1..633
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone="UUGCIM0354H14"
/clone_1ib="Mouse 10kb plasmid UUGCIM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57Bl/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|g91AF12072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 147 a 154 c 134 g 198 t
ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 707 tggcacatcagaagcaga 726
|||||
Db 257 TGGCACAATCAGAGACAGA 276

RESULT 70
AM874886 644 bp mRNA EST 22-MAY-2000
LOCUS SMOVNFCA10G04SK Onchocerca volvulus microfilaria cDNA
DEFINITION (SAM98MLM-OVMF) Onchocerca volvulus cDNA clone SMOVNFCA10G04 5', mRNA sequence.
ACCESSION AM874886
VERSION AM874886.1 GI:8012581
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
1 (bases 1 to 644)
Williams, S.A.
Genes expressed in microfilaria of *Onchocerca volvulus*
Unpublished (1999)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pbluescript SK.

JOURNAL COMMENT

FEATURES
Source

location/Qualifiers
1.644
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOV3MCA07B09"
/clone_lib="Onchocerca volvulus microfilaria cDNA (SAM98MLM-Ovml3)"
/dev_stage="microfilaria"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email:genomesmith.edu."

BASE COUNT
182 a 125 c 152 g 185 t

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 644;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgcataaact 2383
|||||
Db 27 GAAGAAGATGCTATTAACACT 46

RESULT 71
BE346038 652 bp mRNA EST 17-JUL-2000
LOCUS SMOV3MCA07B09SK Onchocerca volvulus L2 larvae cDNA (SAM98MLM-Ovml3)
DEFINITION Onchocerca volvulus cDNA clone SMOV3MCA07B09 5', mRNA sequence.
ACCESSION BE346038
VERSION BE346038.1 GI:9255574
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE
1 (bases 1 to 652)
Williams, S.A.
Genes expressed in L2 larvae of Onchocerca volvulus
Unpublished (1999)
COMMENT
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1.652
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOV3MCA07B09"
/clone_lib="Onchocerca volvulus L2 larvae cDNA (SAM98MLM-Ovml3)"
/dev_stage="L2"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 9,000 L2s isolated from infected mosquitoes from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library

has 7.3 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email: genomesmith.edu."

BASE COUNT
187 a 124 c 157 g 184 t

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 652;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2364 gaagaagatgcataaact 2383
|||||
Db 27 GAAGAAGATGCTATTAACACT 46

RESULT 72
A1317901 653 bp mRNA EST 17-DEC-1998
LOCUS SMOV3MCA07B09SK Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)
DEFINITION Onchocerca volvulus cDNA clone SMOV3MCA07B09 5', mRNA sequence.
ACCESSION A1317901
VERSION A1317901.1 GI:4033168
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE
1 (bases 1 to 653)
Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
COMMENT
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1.653
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCA07B09"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mJ3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E5 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigmaybc.org)."

BASE COUNT
184 a 127 c 160 g 182 t

Query Match 0.6%; Score 20; DB 10; Length 653;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383
 Db 31 GAAGAAGATGCTATAAAACT 50

RESULT 73

BE636495

LOCUS BE636495 656 bp mRNA EST 25-AUG-2000
 DEFINITION SMOVL2CAS17D04SK Onchocerca volvulus L2 larvae cDNA (SAW98MLW-OVL2)

ONCHOCERCA VOLVULUS CDNA clone SMOVL2CAS17D04 5', mRNA sequence.

BE636495

BE636495.1 GI:9919606

KEYWORDS

EST.

SOURCE Onchocerca volvulus.

ORGANISM Onchocerca volvulus

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 656)

AUTHORS Williams, S.A.

TITLE Genes expressed in L2 larvae of Onchocerca volvulus

JOURNAL Unpublished (1999)

COMMENT Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pBluescript SK.

Location/Qualifiers

1. 656

/organism="Onchocerca volvulus"

/db_xref="taxon:6282"

/clone="SMOVL2CAS17D04"

/clone_1lb="Onchocerca volvulus L2 larvae cDNA

(SAW98MLW-OVL2)"

/dev_stage="L2"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-Zap XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. mRNA was

prepared from approximately 9,000 L2s isolated from

infected mosquitoes from Kumba, Cameroon and converted to

double-stranded cDNA using reverse transcriptase and

oligo(dT) followed by RNase H and DNA pol I. The library

has 7.3 x 10B4 independent recombinants and the average

insert size is approximately 1kb. The library was

constructed by Michelle Lizotte-Waniewski. The library is

available from Dr. S.A. Williams, email: genome@smith.edu."

BASE COUNT 183 a 129 c 157 g 186 t 1 others

ORIGIN

Query Match 0.6%; Score 20; DB 10; Length 656;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383
 Db 28 GAAGAAGATGCTATAAAACT 47

RESULT 74

BG310513

LOCUS BG310513 660 bp mRNA EST 23-FEB-2001

DEFINITION SMOV3MCAM56C02SK Onchocerca volvulus molting L3 larva cDNA

(SL96MLW-OvML3) Onchocerca volvulus cDNA clone SMOV3MCAM56C02 5',

mRNA sequence.

ACCESSION BG310513
 VERSION BG310513.1 GI:13112371
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 660)

AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.

TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus

JOURNAL Unpublished (1997)

COMMENT Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: genome@smith.edu

Seq primer: pBluescript SK.

Location/Qualifiers

1. 660

/organism="Onchocerca volvulus"

/strain="Kumba, Cameroons"

/db_xref="taxon:6282"

/clone="SMOV3MCAM56C02"

/clone_1lb="Onchocerca volvulus molting L3 larva cDNA

(SL96MLW-OvML3)"

/dev_stage="molting L3"

/lab_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-Zap XR; Site_1: Eco RI; Site_2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in

Cameroon (forest strain). The L3 were cultured in 20% FCS

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in

culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately

6000 molting larvae (mL3). 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using

reverse transcriptase and oligo(dT) followed by RNase H

and DNA pol I. The library was constructed in the lambda

Uni-Zap XR vector and has 1 x 10B6 independent

recombinants and the average insert size is ~1200 bp. The

library was constructed by Sara Lustigman and Michelle

Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email:

slustig@nhydc.org)."

BASE COUNT 187 a 127 c 161 g 185 t

ORIGIN

Query Match 0.6%; Score 20; DB 11; Length 660;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2364 gaagaagatgctataaact 2383
 Db 28 GAAGAAGATGCTATAAAACT 47

RESULT 75

CNS02935/c

LOCUS CNS02935/c

DEFINITION

Tetradon nigroviridis genome survey sequence PUC-01 end of clone

247A05 of library G from Tetradon nigroviridis, genomic survey

sequence.

ACCESSION AL186746

VERSION AL186746.1 GI:7824850

KEYWORDS GSS: genome survey sequence.

SOURCE Tetradon nigroviridis.

ORGANISM Tetradon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

REFERENCE
AUTHORS

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

1 (bases 1 to 726)

Roeest-Croillius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

Characterization and repeat analysis of the compact genome of the
freshwater puterfish Tetraodon nigroviridis

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished

2 (bases 1 to 726)

Roeest-Croillius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 726)

TITLE
JOURNAL
REFERENCE
AUTHORS
COMMENT

Genoscope.

Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at

<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
SOURCE

Location/Qualifiers

1..726

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="247A05"

/clone_lib="G"

/note="Genoscope sequence ID : C0AG247A03SP1-end :

PUC-Or1"

BASE COUNT 231 a 162 c 139 g 185 t 9 others

ORIGIN

Query Match 0.6%; Score 20; DB 13; Length 726;

Best Local Similarity 100.0%; Pred.No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 215 aaaggttcagagtcctgta 234

|||||

DB 533 AAAGGTTTCAGAGTCCTGTA 514

Search completed: March 25, 2002, 15:27:50
Job time: 13639 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:39:43 ; Search time 97.58 Seconds

(without alignments)
7129.949 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcctaagacacaa.....ctttaactagtaactgct 3072

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0
351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	1.9	5502	3	US-08-836-134-1 Sequence 1, Appl
2	39.6	1.3	2862	4	US-08-569-749-13 Sequence 13, Appl
3	39.6	1.3	2862	5	PCR-US96-12860-13 Sequence 13, Appl
4	39.6	1.3	3151	3	US-09-212-971-13 Sequence 13, Appl
5	39.6	1.3	3151	3	US-08-800-929A-13 Sequence 13, Appl
6	39.6	1.3	3151	4	US-09-617-053A-13 Sequence 13, Appl
7	36.8	1.2	1435	4	PCR-US95-05922A-1 Sequence 1, Appl
8	36.8	1.2	2589	4	US-08-569-749-1 Sequence 1, Appl
9	36.8	1.2	2589	5	PCR-US96-12860-1 Sequence 1, Appl
10	36.8	1.2	3532	2	US-09-205-204-1 Sequence 1, Appl
11	36.8	1.2	3732	3	US-09-212-971-7 Sequence 7, Appl
12	36.8	1.2	3732	3	US-08-800-929A-7 Sequence 7, Appl
13	36.8	1.2	3732	4	US-08-617-053A-7 Sequence 7, Appl
14	35.2	1.1	2580	2	US-08-511-485-7 Sequence 7, Appl
15	34.8	1.1	1840	2	US-08-484-993B-15 Sequence 15, Appl
16	34.8	1.1	1840	2	US-08-484-156B-15 Sequence 15, Appl
17	34.8	1.1	1840	2	US-08-484-596A-15 Sequence 15, Appl
18	34.8	1.1	1840	2	US-08-480-150A-15 Sequence 15, Appl
19	34.8	1.1	1840	3	US-08-458-731-15 Sequence 15, Appl
20	34.8	1.1	1840	3	US-08-149-223A-15 Sequence 15, Appl
21	34.8	1.1	53526	3	US-08-658-136-2 Sequence 2, Appl
22	34.8	1.1	53577	3	US-08-658-136-1 Sequence 2, Appl
23	34.2	1.1	510	2	US-08-934-959-7 Sequence 7, Appl
24	34.2	1.1	7218	1	US-08-232-463-14 Sequence 14, Appl
25	34	1.1	200	2	US-08-875-972-27 Patent No. 525348
26	34	1.1	1753	6	525348-2 Patent No. 525348
27	34	1.1	4695	6	525348-3 Patent No. 525348

28	33.8	1.1	292	1	US-08-644-664B-41 Sequence 41, Appl
29	33.8	1.1	292	2	US-08-761-277A-41 Sequence 41, Appl
30	33.8	1.1	3979	4	US-09-180-439-1 Sequence 1, Appl
31	33.8	1.1	3979	4	US-09-180-439-2 Sequence 2, Appl
32	33.8	1.1	4123	4	US-09-180-439-7 Sequence 7, Appl
33	33.4	1.1	893	4	US-09-142-320-3 Sequence 3, Appl
34	33.4	1.1	893	4	US-09-142-320-8 Sequence 8, Appl
35	33.4	1.1	1359	2	US-08-851-088-1 Sequence 1, Appl
36	33.4	1.1	3573	4	US-09-353-585-4 Sequence 4, Appl
37	33.4	1.1	5535	1	US-08-089-755A-1 Sequence 1, Appl
38	33.4	1.1	5535	1	US-08-089-755A-4 Sequence 4, Appl
39	33.4	1.1	5535	1	US-08-421-754-1 Sequence 1, Appl
40	33.4	1.1	5535	1	US-08-421-754-4 Sequence 4, Appl
41	33.4	1.1	5535	2	US-08-421-791-1 Sequence 1, Appl
42	33.4	1.1	5535	2	US-08-421-791-4 Sequence 4, Appl
43	33.4	1.1	6471	4	US-09-353-585-1 Sequence 1, Appl
44	33.2	1.1	3624	1	US-07-951-715A-6 Sequence 6, Appl
45	33.2	1.1	3624	2	US-08-459-448A-6 Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-836-134-1
; Sequence 1, Application US/08836134A
; Patent No. 6020127
; GENERAL INFORMATION:
; APPLICANT: Mackenzie, Alex E.
; APPLICANT: Korneluk, Robert G.
; APPLICANT: Mahdevan, Mani S.
; APPLICANT: Mclean, Michael
; APPLICANT: Roy, Natalie
; APPLICANT: Ikeda, John
; TITLE OF INVENTION: Neuronal Apoptosis Inhibitor Protein, Gene Sequence and Patent No. 6020127
; TITLE OF INVENTION: Mutations Causative of Spinal Muscular Atrophy
; FILE REFERENCE: 3477-112, 033477/139914
; CURRENT APPLICATION NUMBER: US/08/836,134A
; CURRENT FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5502
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-836-134-1

Query Match 1.9%; Score 59; DB 3; Length 5502;
Best Local Similarity 54.3%; Pred. No. 3.3e-07;
Matches 144; Conservative 0; Mismatches 115; Indels 6; Gaps 1;
QY 500 tbaaggggaatctggcaaggaagcaagtcacactgtctgcgcgcgcatgcatgctggtg 559
DB 1798 tggaggggtgaagctgtgaagtggaagacggtctctcttggaagaatagcttctgtgg 1857
QY 560 gctcgcgaagtgcaaggtctgacaaagttcaatctgtctctctctcgtcgtcgaag 617
DB 1858 catctgcatgctgtcccccgttaaacaggtlccaggtgtttcttaccctcccttaagt 1917
QY 618 ----caggcccgaggtgtaacttttgaacacctctgtatacaacctcgtgataccg 673
DB 1918 ccaccagaccagcagggaggtctgcccagatcatctgttaccacgcctcagagaagaag 1977
QY 674 gcaacatcggaaagacatcatgcatgctgtgtgaagctgcgcgcgagaggtctt 733
DB 1978 gatctgttactgaaatgctcatgaggaacatcatccagcagttaaagaatcaggtctat 2037
QY 734 tccctcttgagctacaatgaatt 758
DB 2038 tccctttagatgactacaagaat 2062

```
RESULT 2
US-08-569-749-13
; Sequence 13, Application US/08569749
; Patent No. 6187557
; GENERAL INFORMATION:
; APPLICANT: Rothe, Mike
; APPLICANT: Goeddel, David V
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,749
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J.
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: A-62464/DJB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)781-1989
; TELEFAX: (415)398-3249
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2862 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-569-749-13

Query Match 1.3%; Score 39.6; DB 4; Length 2862;
Best Local Similarity 46.9%; Pred. No. 0.15;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 50 tgactgtataaagcaatacagatgactatttgatggaatgctcgaatcggaag 109
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1869 TGACACATGTCCTTCCTGATCTCTTGTGAGGCCAGTGTAATTACAAACAGG 1928
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 aagtaaacatcattgtcgagaagtgagcagatgctgctcagaaggatcaca 169
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1929 AACATGATATTATTATAGACAGAAAAACACAGATACCTTACAGCAGAGAGCTTTTGACA 1988
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 tgaattgaaaagggtcagagtcctgtaacctcttcttaaalcccttaaggagtga 229
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1989 CCGTTTATAGTCAAGGAAATGCTGCAGCCAAATCTCTCAAAAACCTCTCTGAAGAAATTG 2048
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 230 actatcctatttcagagacttgatgacaagaatcttttcacagacatcagaagag 289
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2049 ACTGCACGTTATATGAATACTTATTGTGAGAAAAGAAATATGTAATTCACACAGAAG 2108
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 290 actgagcagattggctcagga 311
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2109 ACGTTTCAGGCTTGTCATTGGA 2130
    || | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
PCT-US96-12860-13
; Sequence 13, Application PC/TUS9612860
; GENERAL INFORMATION:
; APPLICANT: TULARIK, INC.
```

```
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/12860
; FILING DATE: 06 AUG 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Serial Nos. 08/512,946 & 08/569,749
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J.
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: A-62464/DJB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)781-1989
; TELEFAX: (415)398-3249
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2862 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US96-12860-13

Query Match 1.3%; Score 39.6; DB 5; Length 2862;
Best Local Similarity 46.9%; Pred. No. 0.15;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 50 tgactgtataaagcaatacagatgactatttgatggaatgctcgaatcggaag 109
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1869 TGACACATGTCCTTCCTGATCTCTTGTGAGGCCAGTGTAATTACAAACAGG 1928
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 aagtaaacatcattgtcgagaagtgagcagatgctgctcagaaggatcaca 169
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1929 AACATGATATTATTATAGACAGAAAAACACAGATACCTTACAGCAGAGAGCTTTATGACA 1988
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 tgaattgaaaagggtcagagtcctgtaacctcttcttaaalcccttaaggagtga 229
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1989 CCGTTTATAGTCAAGGAAATGCTGCAGCCAAATCTCTCAAAAACCTCTCTGAAGAAATTG 2048
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 230 actatcctatttcagagacttgatgacaagaatcttttcacagacatcagaagag 289
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2049 ACTGCACGTTATATGAATACTTATTGTGAGAAAAGAAATATGTAATTCACACAGAAG 2108
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 290 actgagcagattggctcagga 311
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2109 ACGTTTCAGGCTTGTCATTGGA 2130
    || | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-09-212-971-13
; Sequence 13, Application US/09212971B
; GENERAL INFORMATION:
; APPLICANT: Korneljuk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
```


SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13
LENGTH: 3151
TYPE: DNA
ORGANISM: Mus musculus
US-09-617-053A-13

Query Match 1.3%; Score 39.6; DB 4; Length 3151;
Best Local Similarity 46.9%; Pred. No. 0.16;
Matches 123; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 50 tgaactgtataaagaacacacagatgacatttggatggaatgttcgtaacgcggaag 109
DB 2175 tgaacatgtctctccatcccgatgatactctcttgagcgcaagtgaattacaacag 2234
QY 110 aagtaaacctcttctgctcggaaggtgagcagaatgctgctgaagatcattcaca 169
DB 2235 aacatgatatattatgacagaacacacagatcccttaacagaagagctattgaca 2294
QY 170 tgatttgaagaaggttcagagctgtaacctcttcttaaccccttaaggaagtga 229
DB 2295 ccgttttgaatcagggaatgtctgcagccaacatcttcaaaaactctcgaaggaatg 2354
QY 230 actatccttatttcagagactgaaatggaacaaagcttcttcaatcagacatcagaag 289
DB 2355 actcaagtgatatagaacattatttggaaaagataatgatatattccacagaag 2414
QY 290 acttggacgatttgctcaga 311
DB 2415 acgttcaggcttgatcattga 2436

RESULT 7

PCT-US95-05922A-1

Sequence 1, Application PC/TUS9505922A

GENERAL INFORMATION:

APPLICANT: HE, ET AL.

TITLE OF INVENTION: Human Inhibitor of Apoptosis Gene 1

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

ADDRESS: CECCHI, STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

CITY: ROSELAND

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/05922A

FILING DATE: 11 MAY 1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: FERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 325800-292

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1435 BASE PAIRS

TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR

MOLECULE TYPE: CDNA
PCT-US95-05922A-1

Query Match 1.2%; Score 36.8; DB 5; Length 1435;
Best Local Similarity 46.2%; Pred. No. 0.69;
Matches 122; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 48 aagactgtatataaagaacatcacagatgacatttggatggaatgttcgaatgcga 107
DB 882 ATTGACATGTGTCTCTTCCTATCTCGATATATCTTTTAAAGCCCAATGATTAATAACA 941
QY 108 agaagtaaacatcatttgcgcgaaggtgagcagaatgctgctagaagggatcaca 167
DB 942 GGNACATGATATATTATTAACAAAACACAGATACCTTTACAGCGACGAGAACTGATGA 1001
QY 168 catgatttgaagaaggttcagagctcgtgaacctcttcttaaatcccttaaggaagt 227
DB 1002 TACCATTGTTGGTTAAAGAAATGCTCGGCCAACATCTTCMAAAACTGCTTAAAGAAAT 1061
QY 228 gaactatcctctatcttcagagactgaaatgacaaagcttcttcatcagacacagaag 287
DB 1062 TGACTCTACATTGTATTAAGACTTATTGTGATGAATGATATGAACTATATTCCACAGA 1121
QY 288 agacttggacgatttgctcaga 311
DB 1122 AGATGTTTCAGGTCTGTCACACTGA 1145

RESULT 8

US-08-569-749-1

Sequence 1, Application US/08569749

Patent No. 6187557

GENERAL INFORMATION:

APPLICANT: Rothe, Mike

APPLICANT: Goeddel, David V

TITLE OF INVENTION: INHIBITORS OF APOPTOSIS

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/569,749

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Brezner, David J.

REGISTRATION NUMBER: 24,774

REFERENCE/DOCKET NUMBER: A-62464/DJB

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415)781-1989

TELEFAX: (415)398-3249

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2589 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-569-749-1

Query Match 1.2%; Score 36.8; DB 4; Length 2589;
Best Local Similarity 46.2%; Pred. No. 1;

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1840 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Felis domesticus
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 57..1766
US-08-484-993B-15

Query Match 1.1%; Score 34.8; DB 2; Length 1840;
Best Local Similarity 48.5%; Pred. No. 3.3;
Matches 127; Conservative 0; Mismatches 132; Indels 3; Gaps 1;
2628 cgtgctagaacagctaccgcagctgctgcccctgggctgtgacgtgcaagcagcct 2687
1604 CTGTGCTAGAGATGCTGACGTACTGCTGTGATGATGAGGTGAGGTGAGGTGCTTCTGCGCT 1545
QY 2688 g---agcagcctgttgaacatttgagagaggtcccaaacctcgtaagcttgggtgaa 2744
1544 GGCACACAGACAGCTTATCTTACAGAGAGAGGTCCACAGCGCTGCGCATGATACATT 1485
QY 2745 aaactggagactcacagatcacagagattagaatttagtgtcatlitttggaaagaaacc 2804
1484 ACAGTGCAGATACCGGTCCTGAGTGCCACTTTGCCATGATGTGCCAAGCTGAA 1425
QY 2805 tctgaaaacttccagcagltgaatttgcggaatcgttgagcggtgatlgtatgct 2864
1424 GGTGAATAACTGAGCGCTTGTAGTAAGATGGAATGATGATCCAGAGCCCTTGAGAC 1365
QY 2865 tgcctcatgggtgtatttgag 2886
1364 AGGATCAGCTGGGTTGATAG 1343

Search completed: March 25, 2002, 11:40:28
Job time: 3012 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:44:40 ; Search time 270.82 Seconds

(without alignments)
9724.921 Million cell updates/sec

Title: US-09-697-089-3

Perfect score: 3072

Sequence: 1 atgaattcataaaggacaa.....ctttaactagtaactgct 3072

Printing table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3072	100.0	3133	22	AA503945	Human caspase recr
2	3067.2	99.8	3545	22	AAH98254	Murine EST-derived
3	2980.6	97.0	3615	22	AAH98254	Human caspase recr
4	1920.6	62.5	2950	22	AAH99581	Human protein enco
5	538.4	17.5	2735	22	AAH34171	Human colon cancer
6	242	7.9	421	22	AA114389	Probe #4322 for 9e
7	242	7.9	421	22	AA135764	Probe #4450 used t
8	242	7.9	421	22	AA104213	Probe #4204 used t
9	220	7.2	220	22	AA123590	Probe #13523 for 9
10	220	7.2	220	22	AA148904	Probe #17590 used
11	220	7.2	220	22	AA109206	Probe #9197 used t

12	91.8	3.0	936	22	AAF58252	Oligonucleotide D1
13	91.8	3.0	936	22	AAF58254	Oligonucleotide D1
14	91.8	3.0	936	22	AAF58257	Oligonucleotide D1
15	91.8	3.0	936	22	AAF58259	Oligonucleotide D2
16	91.8	3.0	936	22	AAF58262	Oligonucleotide D2
17	91.8	3.0	938	22	AAF58255	Oligonucleotide D1
18	91.4	3.0	936	22	AAF58252	Oligonucleotide D1
19	91.4	3.0	936	22	AAF58254	Oligonucleotide D1
20	91.4	3.0	936	22	AAF58257	Oligonucleotide D1
21	91.4	3.0	936	22	AAF58259	Oligonucleotide D2
22	91.4	3.0	936	22	AAF58262	Oligonucleotide D2
23	91.4	3.0	938	22	AAF58255	Oligonucleotide D1
24	59	1.9	5366	20	AA58001	Conodontropic hormo
25	59	1.9	5366	20	AA58273	Human apoptosis in
26	59	1.9	5502	17	AA730092	Neuronal apoptosis
27	59	1.9	5502	18	AA71263	Neuronal apoptosis
28	59	1.9	5984	20	AA58000	Conodontropic hormo
29	59	1.9	5984	20	AA58272	Human apoptosis in
30	59	1.9	5984	21	AA39808	Human Nalp CDNA.
31	59	1.9	6124	18	AA71265	Neuronal apoptosis
32	59	1.9	6133	18	AA71264	Neuronal apoptosis
33	59	1.9	6228	18	AA71266	Neuronal apoptosis
34	42	1.4	244	22	AAF58238	Oligonucleotide D1
35	41.4	1.3	4141	20	AA209251	Murine CARD-4L cDN
36	41.4	1.3	4141	22	AA30006	Mouse CARD-4L (lon
37	39.8	1.3	3080	20	AA209248	Human CARD-4S part
38	39.8	1.3	3080	22	AA30003	Human CARD-4S (sho
39	39.8	1.3	3382	22	AA30002	Human CARD-4L (lon
40	39.8	1.3	4415	21	AAA95791	Apoptosis related
41	39.8	1.3	4610	22	AAH18313	Human CDNA sequenc
42	39.8	1.3	32042	22	AA209252	Human CARD-4 DNA.
43	39.8	1.3	32042	22	AA30011	Human CARD-4 gene.
44	39.6	1.3	244	22	AAF58238	Oligonucleotide D1
45	39.6	1.3	2862	18	AA761592	Murine C-IAP. Mus

ALIGNMENTS

RESULT 1	
AA503945	
ID	AA503945 standard; cDNA; 3133 BP.
XX	
AC	AA503945;
DT	12-SEP-2001 (first entry)
XX	
DE	Human caspase recruitment domain 12 (CARD-12) cDNA.
XX	
KW	Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
KW	cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
KW	systemic lupus erythematosus; arthritis; neurological disorder; stroke;
KW	Alzheimer's disease; amyotrophic lateral sclerosis; haematologic disease;
KW	aplastic anaemia; myocardial infarction; inflammatory disorder;
KW	Crohn's disease; insulin-dependent diabetes; contact dermatitis;
KW	psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
KW	tuberculosis; ischaemic brain injury; hypoxic brain injury; ss;
KW	kidney ischaemia; reperfusion injury; acute bacterial meningitis;
KW	excitotoxic brain damage; liver disease.
OS	Homo sapiens.
XX	
EH	Key
FT	Location/Qualifiers
FT	36..3110
FT	/*tag= a
FT	/product= "Human CARD-12"
PN	WO200130971-A2.
XX	
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-US29643.
XX	

PR 27-OCT-1999; 99US-0161822.
XX (MILL-) MILLENNIUM PHARM INC.
XX
PI Berlin J, Robison KE;
XX WPI: 2001-308628/32.
DR P-PSDB: AAD02880.
XX
XX Isolated caspase recruitment domain-12 polypeptide and nucleic acids
PT encoding them, useful for treating and diagnosing disorders associated
PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
PT disease -
XX
XX Claim 2; Fig 1; 93pp; English.
XX
XX The sequence represents a cDNA which encodes the human caspase
CC recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
CC number of proteins that transmit signals that activate apoptosis and in a
CC inflammatory pathways in response to stress and other stimuli. Therefore,
CC CARD-12 and its corresponding nucleic acid may be used in treatment and
CC diagnosis of patients suffering from disorders associated with an
CC abnormal level (an increase or a decrease) of apoptotic cell death or
CC abnormal activity of stress-related pathways. The disorders include
CC cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
CC autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
CC neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
CC sclerosis), haematologic diseases (e.g. aplastic anaemia, myocardial
CC infarction, stroke), inflammatory and immune system disorders (e.g.
CC Crohn's disease, insulin-dependent diabetes, contact dermatitis,
CC psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
CC leptomonous leprosy), ischemic and hypoxic brain injury, kidney
CC ischemia/reperfusion injury, excitotoxic brain damage, acute bacterial
CC meningitis and liver disease.
XX
XX Sequence 3133 BP; 903 A; 691 C; 729 G; 810 T; 0 other:
SQ

Query Match 100.0%; Score 3072; DB 22; Length 3133;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3072; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaattcataaagacatagccgagcccttatcaagaatggatgactgtata 60
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 36 atgaattcataaagacatagccgagcccttatcaagaatggatgactgtata 95
QY 61 aagcaatcacagatgactatttgatgaaatgtctgaaacgcaagaataaacatc 120
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 96 aagcaatcacagatgactatttgatgaaatgtctgaaacgcaagaataaacatc 155
QY 121 attgctcgagaaagtgagcagagatgctctagagagatcatcatgatttga 180
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 156 atttctcgagaaagtgagcagagatgctctagagagatcatcatgatttga 215
QY 181 aaggtttagagctctgttaaccttttctaacttaacgagatgagactatctcta 240
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 216 aaggtttagagctctgttaaccttttctaacttaacgagatgagactatctcta 275
QY 241 ttccagagactgaaatgagcaaatcttttcatcagacatcagaaggagacttgagat 300
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 276 ttccagagactgaaatgagcaaatcttttcatcagacatcagaaggagacttgagat 335
QY 301 ttggtcagagattaaagactgttacatcccatctttctgaaactttatccctt 360
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 336 ttggtcagagattaaagactgttacatcccatctttctgaaactttatccctt 395
QY 361 ggtgaagatattgacatatattttaacttgaaaagcacttcaacgaaccttctgtg 420
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 396 ggtgaagatattgacatatattttaacttgaaaagcacttcaacgaaccttctgtg 455
QY 421 aggaagaccaaaccatcacccgctgagcagctgacccctgaatggcctctcgagct 480
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 456 aggaagaccaaaccatcacccgctgagcagctgacccctgaatggcctctcgagct 515

QY 481 ctccagagccctgacatcatgaaaggaaatctcgcaaaagcgaagctcgtctgag 540
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 516 ctccagagccctgacatcatgaaaggaaatctcgcaaaagcgaagctcgtctgag 575
QY 541 cgcattgccaatgctctcgaggctccggaagaatgcgaagctctgacaaagtccaatctc 600
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 576 cgcattgccaatgctctcgaggctccggaagaatgcgaagctctgacaaagtccaatctc 635
QY 601 ttctctcgtctcgaagcgaagcccggtgagcttttgaacccctctgtacaatc 660
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 636 ttctctcgtctcgaagcgaagcccggtgagcttttgaacccctctgtacaatc 695
QY 661 ctggatatccctgacaaatcaagaagcagacatcatgagccatgctgctgaagctgcg 720
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 696 ctggatatccctgacaaatcaagaagcagacatcatgagccatgctgctgaagctgcg 755
QY 721 caaaggtctcttctctcttctgagctgaacaaatgaattcaagcccaagactccagaa 780
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 756 caaaggtctcttctctcttctgagctgaacaaatgaattcaagcccaagactccagaa 815
QY 781 atcgaagccctgataaagaaacacccgcttcaagaacatggtcatgctgacacatc 840
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 816 atcgaagccctgataaagaaacacccgcttcaagaacatggtcatgctgacacatc 875
QY 841 actgaatgctcgaagcacaatcagcagtttgtgctcctgactgctgaagtggggatag 900
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 876 actgaatgctcgaagcacaatcagcagtttgtgctcctgactgctgaagtggggatag 935
QY 901 acgaagaacacgcccagctctcctacccgaagaatgctgatacgaagatctgtgaagc 960
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 936 acgaagaacacgcccagctctcctacccgaagaatgctgatacgaagatctgtgaagc 995
QY 961 ttgtgctcacaatcagaatccagtgctgtgagaaatctcatgagaagccctctctt 1020
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 996 ttgtgctcacaatcagaatccagtgctgtgagaaatctcatgagaagccctctctt 1055
QY 1021 gtgtgtacactgtgcaatccagatggtgtaagaatgagttccatctcacacacaa 1080
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1056 gtgtgtacactgtgcaatccagatggtgtaagaatgagttccatctcacacacaa 1115
QY 1081 agcgtgtccatccctctcatgctgtgtgatacagaataaacaaacaaacaaatgaagt 1140
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1116 agcgtgtccatccctctcatgctgtgtgatacagaataaacaaacaaacaaatgaagt 1175
QY 1141 gtgtgtcgaagtacttcatctcgaagcctggaacactgtggaacactagctcgaaggt 1200
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1176 gtgtgtcgaagtacttcatctcgaagcctggaacactgtggaacactagctcgaaggt 1235
QY 1201 gtgtgtcgaagtacttcatctcgaagcctggaacactgtggaacactagctcgaaggt 1260
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1236 gtgtgtcgaagtacttcatctcgaagcctggaacactgtggaacactagctcgaaggt 1295
QY 1261 ctgtgtaacaatggtgctctctgttaatatcagcctcaaggttcaagccaaagtataa 1320
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1296 ctgtgtaacaatggtgctctctgttaatatcagcctcaaggttcaagccaaagtataa 1355
QY 1321 ttctttcacaagatcatctcagaagatcacacagcagaagactcagacttattgagc 1380
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1356 ttctttcacaagatcatctcagaagatcacacagcagaagactcagacttattgagc 1415
QY 1381 tctcatgagcgaagagatgaccaaagggaatggttacttgcagaanaatggtttccatt 1440
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1416 tctcatgagcgaagagatgaccaaagggaatggttacttgcagaanaatggtttccatt 1475
QY 1441 tgggaataacatccactatagcagcgtctccgtgtacacctgtgtgcatcttgagaa 1500
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1476 tgggaataacatccactatagcagcgtctccgtgtgtacacctgtgtgcatcttgagaa 1535
QY 1501 gccacaggcgtgttataagacccctcgagcaggttatcaaacagcgtcctctcgaa 1560
DB |||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1536 gccacaggcgtgttataagacccctcgagcaggttatcaaacagcgtcctctcgaa 1595

1561 cttccacgcgaagaggcctctctgagacagaatcttgcagaagtgcgaacacc 1620
1596 cttccacgcgaagaggcctctctgagacagaatcttgcagaagtgcgaacacc 1655
1621 actgagcaagaattctgaaagccataacatcaatcccttctgtagagtgcacat 1680
1656 actgagcaagaattctgaaagccataacatcaatcccttctgtagagtgcacat 1715
1681 ttatctcagaagagatctcacaatccagccctgagccaagaatttggaagcttcttcaa 1740
1716 ttatctcagaagagatctcacaatccagccctgagccaagaatttggaagcttcttcaa 1775
1741 ggtcagaagcttatataataactcagaaggaaaccccgattacttattgactcttgcga 1800
1776 ggtcagaagcttatataataactcagaaggaaaccccgattacttattgactcttgcga 1835
1801 catttgcccaattctgcaagtgctctggaactcattaaactggaactttatgaggagct 1860
1836 catttgcccaattctgcaagtgctctggaactcattaaactggaactttatgaggagct 1895
1861 atggtttatgaggaaagagctgagcaagacagctggaatccacatggaagagcccca 1920
1896 atggtttatgaggaaagagctgagcaagacagctggaatccacatggaagagcccca 1955
1921 gaaacctacatcccgagcagagctgtatcttcttctcgaactggaagcagaatccag 1980
1956 gaaacctacatcccgagcagagctgtatcttcttctcgaactggaagcagaatccag 2015
1981 actctgagagtcacaccccgagatttcagcaagttgaaatgaagaatatcacatctg 2040
2016 actctgagagtcacaccccgagatttcagcaagttgaaatgaagaatatcacatctg 2075
2041 ggggaaatatcagctgcgcacaaagcctcaggtcgcgaataaagaagatgctggtg 2100
2076 ggggaaatatcagctgcgcacaaagcctcaggtcgcgaataaagaagatgctggtg 2135
2101 gctggaagcctcaggttgcctcagcagcctgtaagaacattatctctcagtgtaa 2160
2136 gctggaagcctcaggttgcctcagcagcctgtaagaacattatctctcagtgtaa 2195
2161 gccagtcctccccaatgaaagatgagagacatcacatctgttaacaaacctgaaacc 2220
2196 gccagtcctccccaatgaaagatgagagacatcacatctgttaacaaacctgaaacc 2255
2221 ttgagtatcatgactacagaatcaacgctgcgcgggtgctctcagcagacgcttggt 2280
2256 ttgagtatcatgactacagaatcaacgctgcgcgggtgctctcagcagacgcttggt 2315
2281 aactggaagacactatacaagctcataatgataaagaatgaagaagatgct 2340
2316 aactggaagacactatacaagctcataatgataaagaatgaagaagatgct 2375
2341 ataaactagctgaagcctgtaaaaaacctgaaagaatgctttaaattcattggccac 2400
2376 ataaactagctgaagcctgtaaaaaacctgaaagaatgctttaaattcattggccac 2435
2401 ttgctgacatgtagagaggaatgataatcaatgctcctgctcgaagtgaacccgt 2460
2436 ttgctgacatgtagagaggaatgataatcaatgctcctgctcgaagtgaacccgt 2495
2461 gacctggaagaatcaatgactcctcctgctgctgtctgcaaatgagcagtgaaacctta 2520
2496 gacctggaagaatcaatgactcctcctgctgctgtctgcaaatgagcagtgaaacctta 2555
2521 gctcagaatcttcacaatttgcgaactgagcattctgtatctacagaagaattacat 2580
2556 gctcagaatcttcacaatttgcgaactgagcattctgtatctacagaagaattacat 2615
2581 gaaagaagatgaaatgagcctctcattatgaactgtagcagaagatgaaagctgtagaag 2640
2616 gaaagaagatgaaatgagcctctcattatgaactgtagcagaagatgaaagctgtagaag 2675
2641 ctacacgcacgatgctgctcctgagctgctgtagctgcaagcagcctgtagagcctgctg 2700

2676 ctacacgcacgatgctgctcctgagctgctgtagcagcagaagcagcctgtagcctgtg 2735
2701 aaacattgagagaggtccacacactcgtcaagcttggtgtgaaaaactgagactcaca 2760
2736 aaacattgagagaggtccacacactcgtcaagcttggtgtgaaaaactgagactcaca 2795
2761 gatacagaagatgaatttgaatgctcatttttggaaaaaacccctcgaagaactccag 2820
2796 gatacagaagatgaatttgaatgctcatttttggaaaaaacccctcgaagaactccag 2855
2821 cagttgaatttgcgggaatcgtgtgaagcagtgatgtagtgctgccttcagtggtgta 2880
2856 cagttgaatttgcgggaatcgtgtgaagcagtgatgtagtgctgccttcagtggtgta 2915
2881 ttggaatccttaagaacttagtgcttttttgaacttagtactaagaacttactcgtat 2940
2916 ttggaatccttaagaacttagtgcttttttgaacttagtactaagaacttactcgtat 2975
2941 ccagcattagtcagaagaacttagccaagtgatcccaagttacttctcgaagaagct 3000
2976 ccagcattagtcagaagaacttagccaagtgatcccaagttacttctcgaagaagct 3035
3001 aggcctgtgtggtgcaatttgatgtagtgaatccagtggtatccagtgctttaa 3060
3036 aggcctgtgtggtgcaatttgatgtagtgaatccagtggtatccagtgctttaa 3095
3061 ctgataactgct 3072
3096 ctgataactgct 3107

RESULT 2

AAH98254 ID AAH98254 standard; cDNA: 3545 BP.

AC AAH98254;

DT 12-OCT-2001 (first entry)

DE Murine EST-derived coding sequence SEQ ID NO: 111.

KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KW diagnostics; forensic test; gene mapping; genetic disorder;

KW biodiversity; gene therapy; nutrition; ss.

OS Mus musculus.

PN WO200154477-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02687.

PR 25-JAN-2000; 2000US-0491404.

PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V,

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

DR WPI: 2001-476164/51.

DR P-PSDB; AAM23595.

PT Isolated polypeptide for treatment of diseases, diagnostics, raising

PS antibodies and research use -

PS Claim 1; Page 250-251; 1275bp; English.

CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.

XX Sequence 3545 BP; 1038 A; 755 C; 816 G; 936 T; 0 other;

Query Match 99.8%; Score 3067.2; DB 22; Length 3545;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3069; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaattcataaaggacatagccgagcccttatccaagaatgggaatgattata 60
 DB |||||||
 DB 232 atgaattcataaaggacatagccgagcccttatccaagaatgggaatgattata 231
 OY 61 aagcaaatcacagatgacctatttgatgaatgctgcaagaagaataaacatc 120
 DB |||||||
 DB 292 aagcaaatcacagatgacctatttgatgaatgctgcaagaagaataaacatc 351
 OY 121 attgctgcgagaaggtgagcagaagatgctgtagaaggatcatcaatgattgaaa 180
 DB |||||||
 DB 352 attgctgcgagaaggtgagcagaagatgctgtagaaggatcatcaatgattgaaa 411
 OY 181 aagggttcagagctcctgtaacctcttcttaaaccttaagaaggatggaaactctctca 240
 DB |||||||
 DB 412 aagggttcagagctcctgtaacctcttcttaaaccttaagaaggatggaaactctctca 471
 OY 241 ttccagacttgaaatggaacaaagtctttctacacagacatcagaagaagacttgacgat 300
 DB |||||||
 DB 472 ttccagacttgaaatggaacaaagtctttctacacagacatcagaagaagacttgacgat 531
 OY 301 ttggtctcagatgttaagaagactgtgtacatacccaactctttctgtaactttatccctc 360
 DB |||||||
 DB 532 ttggtctcagatgttaagaagactgtgtacatacccaactctttctgtaactttatccctc 591
 OY 361 ggtgaagatgacatacttttaacttgaaaagacaccttcacagaacctgcccgtg 420
 DB |||||||
 DB 592 ggtgaagatgacatacttttaacttgaaaagacaccttcacagaacctgcccgtg 651
 OY 421 aggaagagacaaacacatcacgcgtgagacagctgacacctgaaatggcctctgacgact 480
 DB |||||||
 DB 652 aggaagagacaaacacatcacgcgtgagacagctgacacctgaaatggcctctgacgact 711
 OY 481 ctccagagcccttgatcatcttgaaagggaatctggaagaagcagatccactctgctgcag 540
 DB |||||||
 DB 712 ctccagagcccttgatcatcttgaaagggaatctggaagaagcagatccactctgctgcag 771
 OY 541 cgcattgacatgctctgagggctccggaagaagtgaagctctgacaaagtccaattcgtc 600
 DB |||||||
 DB 772 cgcattgacatgctctgagggctccggaagaagtgaagctctgacaaagtccaattcgtc 831
 OY 601 ttctctctcgtctcaagcagggcccgagggtggacttttgaaaacctctgtgtacatc 660
 DB |||||||
 DB 832 ttctctctcgtctcaagcagggcccgagggtggacttttgaaaacctctgtgtacatc 891
 OY 661 ctggaatactcctggcgaacatcagaagaagacatcatatgacatgctgtctgaagctgg 720
 DB |||||||
 DB 892 ctggaatactcctggcgaacatcagaagaagacatcatatgacatgctgtctgaagctgg 951
 OY 721 cagaaggtctcttctctcttgatgctacaaatgaatccaacccacagaactgcccagaa 780
 DB |||||||
 DB 952 cagaaggtctcttctctcttgatgctacaaatgaatccaacccacagaactgcccagaa 1011
 OY 781 atcgaagccctgtataaagaaacacacgcgttcaagaagaatgctcatctaccactaac 840
 DB |||||||
 DB 1012 atcgaagccctgtataaagaaacacacgcgttcaagaagaatgctcatctaccactaac 1071
 OY 841 actgaagctcctgaagcacatacgcgcagtttggtgcctctgactgctgaagtgaggatcag 900

DB |||||||
 DB 1072 actgaagctcctgaagcacatacgcgcagtttggtgcctctgactgctgagtgaggatcag 1131
 OY 901 aacgaatacagcccccaggtctctatccaggaagtgtgtatcaaggagctgtcgtgaagc 960
 DB |||||||
 DB 1132 aacgaatacagcccccaggtctctatccaggaagtgtgtatcaaggagctgtcgtgaagc 1191
 OY 961 ttgtgtcccaattccagaatccaggtgtcttgagaatctcatgaagaccctctctt 1020
 DB |||||||
 DB 1192 ttgtgtcccaattccagaatccaggtgtcttgagaatctcatgaagaccctctctt 1251
 OY 1021 gtgtgtacactgtgtcaatccagatggtgtgaagtgagtccaacttcacacaaaca 1080
 DB |||||||
 DB 1252 gtgtgtacactgtgtcaatccagatggtgtgaagtgagtccaacttcacacaaaca 1311
 OY 1081 acgctgtccatccctctctatctgtatctgtatcagaagaacaacaacaataaagct 1140
 DB |||||||
 DB 1312 acgctgtccatccctctctatctgtatctgtatcagaagaacaacaacaataaagct 1371
 OY 1141 gtgtgtcagaagtactcatctcggaagcttgagaccactgtgagactagctctggaagt 1200
 DB |||||||
 DB 1372 gtgtgtcagaagtactcatctcggaagcttgagaccactgtgagactagctctggaagt 1431
 OY 1201 gtgtgtcccaagtttgatttcgaactcgaggaatgtgtccagcgtgtaattgagatgtc 1260
 DB |||||||
 DB 1432 gtgtgtcccaagtttgatttcgaactcgaggaatgtgtccagcgtgtaattgagatgtc 1491
 OY 1261 ctgtgacaactgtggccctctgtaaaatatacagctcaaaagttccaagccaaagtataa 1320
 DB |||||||
 DB 1492 ctgtgacaactgtggccctctgtaaaatatacagctcaaaagttccaagccaaagtataa 1551
 OY 1321 ttctttacaagtcattccagagctacacagcaggaagactcaagacttattagc 1380
 DB |||||||
 DB 1552 ttctttacaagtcattccagagctacacagcaggaagactcaagacttattagc 1611
 OY 1381 ttctatgagccagaagaggtgacccaagggaatggttactttgagaataatggttccatt 1440
 DB |||||||
 DB 1612 ttctatgagccagaagaggtgacccaagggaatggttactttgagaataatggttccatt 1671
 OY 1441 tccgacatcatccacttctaaagcctctccgagacccctgtggtcatcctgtgaa 1500
 DB |||||||
 DB 1672 tccgacatcatccacttctaaagcctctccgagacccctgtggtcatcctgtgaa 1731
 OY 1501 gccacaaggcgtttatgaagacacctcgacagcgtgtatcaacaagcgtgctcttcgga 1560
 DB |||||||
 DB 1732 gccacaaggcgtttatgaagacacctcgacagcgtgtatcaacaagcgtgctcttcgga 1791
 OY 1561 ctctcatcgcgaaggcctctctcgagacagaatctttgcaagtggtgaataaacacc 1620
 DB |||||||
 DB 1792 ctctcatcgcgaaggcctctctcgagacagaatctttgcaagtggtgaataaacacc 1851
 OY 1621 actgagcaagaattctgaaagccataacatcaatctcttgtaggtgtgacatcat 1680
 DB |||||||
 DB 1852 actgagcaagaattctgaaagccataacatcaatctcttgtaggtgtgacatcat 1911
 OY 1681 ttatatacagaagatcatcatccaaatcagccctgagccaagaattgaaagcttcttcaa 1740
 DB |||||||
 DB 1912 ttatatacagaagatcatcatccaaatcagccctgagccaagaattgaaagcttcttcaa 1971
 OY 1741 ggtataaagcttatataatcaactcagggaaacatcccgatattctattgactcttggaa 1800
 DB |||||||
 DB 1972 ggtataaagcttatataatcaactcagggaaacatcccgatattctattgactcttggaa 2031
 OY 1801 catttgcccaatgtgcaagtgtctggaacttattaaacttgacttttattgaggagact 1860
 DB |||||||
 DB 2032 catttgcccaatgtgcaagtgtctggaacttattaaacttgacttttattgaggagact 2091
 OY 1861 atggtctcatggtgaaaggtctgagaagacacaggttgaaatccacatggaagagcccca 1920
 DB |||||||
 DB 2092 atggtctcatggtgaaaggtctgagaagacacaggttgaaatccacatggaagagcccca 2151
 OY 1921 gaaactacatcccaagcaggtctgtacttcttcaactcggaagcagaatcag 1980
 DB |||||||

Db 2152 gaaactacattccacagcaggctgtatcttcttctcaactggaagcaggaattccag 2211
 Qy 1981 actctggaggtcacactccgggatttcagcaagtgtgaataagcaagatatacatctgt 2040
 Db 2212 actctggaggtcacactccgggatttcagcaagtgtgaataagcaagatatacatctgt 2271
 Qy 2041 gggaaaatattcacgtctgcacaaagcctcaggtcgtcaataaagaatgtgtgtgtgt 2100
 Db 2272 gggaaaatattcacgtctgcacaaagcctcaggtcgtcaataaagaatgtgtgtgtgt 2331
 Qy 2101 gcttggaaagcctcaggt 2160
 Db 2332 gcttggaaagcctcaggt 2391
 Qy 2161 gccagtcctccctcaccatagaagatgagagggacatcattctgttaacaaactgtgaacc 2220
 2392 gccagtcctccctcaccatagaagatgagagggacatcattctgttaacaaactgtgaacc 2451
 Qy 2221 ttgagttattcagacctacagaatcaacgctgtccgggtgtgtgtgtgtgtgtgtgtgt 2280
 Db 2452 ttgagttattcagacctacagaatcaacgctgtccgggtgtgtgtgtgtgtgtgtgtgt 2511
 Qy 2281 aacttgaagaacctacagaatcaacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2340
 Db 2512 aacttgaagaacctacagaatcaacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2571
 Qy 2341 ataataactgaagcctgt 2400
 Db 2572 ataataactgaagcctgt 2631
 Qy 2401 ttgtctgtcagctgt 2460
 Db 2632 ttgtctgtcagctgt 2691
 Qy 2461 gaccttgaagaatcaactgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2520
 Db 2692 gaccttgaagaatcaactgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2751
 Qy 2521 gctcagaatcttcacaaattgt 2580
 Db 2752 gctcagaatcttcacaaattgt 2811
 Qy 2581 gaaaaagaatgaagaatgaagccttcacaaattgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2640
 Db 2812 gaaaaagaatgaagaatgaagccttcacaaattgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2871
 Qy 2641 ctcaacgcagctgt 2700
 Db 2872 ctcaacgcagctgt 2931
 Qy 2701 aaacatttggggggggt 2760
 Db 2932 aaacatttggggggggt 2991
 Qy 2761 gatacagagattagaatttgggt 2820
 Db 2992 gatacagagattagaatttgggt 3051
 Qy 2821 cagttgaatttggcggaagaatcgt 2880
 Db 3052 cagttgaatttggcggaagaatcgt 3111
 Qy 2881 ttgagaatcttaagcaatttgggt 2940
 Db 3112 ttgagaatcttaagcaatttgggt 3171
 Qy 2941 ccaagattagtcagaagaatcttaagcaatttgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3000
 Db 3172 ccaagattagtcagaagaatcttaagcaatttgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3231
 Qy 3001 aggcctgt 3060
 Db 3232 aggcctgt 3291

Qy 3061 ctgagtaactgct 3072
 Db 3292 ctgagtaactgct 3303

RESULT 3
 AAS03946
 ID AAS03946 standard. DNA: 3615 BP.
 AC AAS03946;
 XX
 XX 12-SEP-2001 (first entry)
 DE Human caspase recruitment domain 12 (CARD-12) genomic DNA.
 XX
 XX Caspase recruitment domain; CARD-12; apoptosis; stress-related pathway;
 KW cancer; viral infection; poxvirus; adenovirus; autoimmune disorder;
 KW systemic lupus erythematosus; arthritis; neurological disorder; stroke;
 KW Alzheimer's disease; amyotrophic lateral sclerosis; hematologic disease;
 KW aplastic anaemia; myocardial infarction; inflammatory disorder;
 KW Crohn's disease; insulin-dependent diabetes; contact dermatitis;
 KW psoriasis; graft rejection; bacterial infection; lepromatous leprosy;
 KW tuberculosis; ischaemic brain injury; hypoxic brain injury; ds;
 KW kidney ischaemia; reperfusion injury; acute bacterial meningitis;
 KW excitotoxic brain damage; liver disease.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 1.3615
 FT CDS /tag= a
 FT /product= "Human CARD-12"
 FT
 XX WO200130971-A2.
 PN
 XX
 PD 03-MAY-2001.
 XX
 XX 26-OCT-2000; 2000WO-US29643.
 PF
 XX 27-OCT-1999; 99US-0161822.
 PR
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 XX Bertin J, Robison KE;
 PI WPI, 2001-308628/32.
 DR P-PSDB; AAU02881.
 DR
 XX
 XX
 PT Isolated caspase recruitment domain-12 polypeptide and nucleic acids
 PT encoding them, useful for treating and diagnosing disorders associated
 PT with abnormal apoptosis such as cancer, arthritis and Alzheimer's
 PT disease -
 PS
 XX
 PS Disclosure; Fig 2; 93pp; English.

The sequence represents a genomic DNA which encodes the human caspase
 recruitment domain 12 (CARD-12) polypeptide. CARD domains are found in a
 number of proteins that transmit signals that activate apoptosis and
 inflammatory pathways in response to stress and other stimuli. Therefore,
 CARD-12 and its corresponding nucleic acid may be used in treatment and
 diagnosis of patients suffering from disorders associated with an
 abnormal level (an increase or a decrease) of apoptotic cell death or
 abnormal activity of stress-related pathways. The disorders include
 cancer, viral infections (e.g. caused by poxviruses, adenoviruses),
 autoimmune disorders (e.g. systemic lupus erythematosus, arthritis),
 neurological disorders (e.g. Alzheimer's disease, amyotrophic lateral
 sclerosis), hematologic diseases (e.g. aplastic anaemia, myocardial
 infarction, stroke), inflammatory and immune system disorders (e.g.
 Crohn's disease, insulin-dependent diabetes, contact dermatitis,
 psoriasis, graft rejection), bacterial infections (e.g. tuberculosis,
 lepromatous leprosy), ischaemic and hypoxic brain injury, kidney
 ischaemia/reperfusion injury, excitotoxic brain damage, acute bacterial

CC meningitis and liver disease.
XX
SQ Sequence 3615 BP; 1041 A; 811 C; 845 G; 918 T; 0 other;

Query Match 97.0%; Score 2980.6; DB 22; Length 3615;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 3047; Conservative 0; Mismatches 4; Indels 54; Gaps 1;

QY 2 tgaattcataagagcataagccagcccttcttcaagaatgggaatgactgtataa 61
DB 419 tgaattctataagagcataagccagcccttcttcaagaatgggaatgactgtataa 478
QY 62 agcaaatcacagatgacctatttgtaatgaaagtcttgaatcgcgagaaatgaacatca 121
DB 479 agcaaatcacagatgacctatttgtaatgaaagtcttgaatcgcgagaaatgaacatca 538
QY 122 ttgtcgcgagaaagtgtagcgagagatgctgtatagaggaatcatcatatgttgaata 161
DB 539 ttgtcgcgagaaagtgtagcgagagatgctgtatagaggaatcatcatatgttgaata 598
QY 182 aggtgtcagagtcctgttaacctcttcttaaatcccttaagaggtgaactatcccttat 241
DB 599 aggtgtcagagtcctgttaacctcttcttaaatcccttaagaggtgaactatcccttat 638
QY 242 ttcaggaacttgaaatggacaaa----- 262
DB 659 ttcaggaacttgaaatggacaaaagtttgagagagacagaattggtctctcttaacatca 718
QY 263 -----gtcttttcatcagacatcagaaagagacttggaaattgtgctc 307
DB 719 cctctctctaataagtccttcttcatcagacagaaagagacttggaaattgtgctc 778
QY 308 aggatttaagagactgtatccatacccatcttcttgaacttlaacccctgtgtgaag 367
DB 779 aggatttaagagactgtatccatacccatcttcttgaacttlaacccctgtgtgaag 838
QY 368 atattgacattatttttaacttgaanaagcaacttcaagaacctgtctgttgaagaaag 427
DB 839 atattgacattatttttaacttgaanaagcaacttcaagaacctgtctgttgaagaaag 898
QY 428 accaaacacatcacgcgtgtgagcagctgaaccttgaatgacctccctgcagaccttaca 467
DB 899 accaaacacatcacgcgtgtgagcagctgaaccttgaatgacctccctgcagaccttaca 938
QY 488 gccctgtcatcttgaaggagaaatctgtgcagaagcagatccactgtctgtcagcgactg 547
DB 959 gccctgtcatcttgaaggagaaatctgtgcagaagcagatccactgtctgtcagcgactg 1018
QY 548 ccattgtctcggggctccggaagaatgcaaggctctgaccaaagtccaattcgtctcttc 607
DB 1019 ccattgtctcggggctccggaagaatgcaaggctctgaccaaagtccaattcgtctcttc 1078
QY 608 tccgtctcagcagggcccgaggtgtgaccttttgaaacccctgtgtatcaactcttgata 667
DB 1079 tccgtctcagcagggcccgaggtgtgaccttttgaaacccctgtgtatcaactcttgata 1138
QY 668 taacctgtgacaatlcagagaaagcagacatlcattggccaatgtctgtgaagcttgagcagaag 727
DB 1139 taacctgtgacaatlcagagaaagcagacatlcattggccaatgtctgtgaagcttgagcagaag 1198
QY 728 ttcttttctcttctgtatgctacaatgaattcaagccccaagaactgtccagaatacgaag 767
DB 1199 ttcttttctcttctgtatgctacaatgaattcaagccccaagaactgtccagaatacgaag 1258
QY 788 ccctgtataaagaaaaccccgcttcaagaacatgtgtcatcgtccacactaccactgagt 847
DB 1259 ccctgtataaagaaaaccccgcttcaagaacatgtgtcatcgtccacactaccactgagt 1318
QY 848 gccctgaagcacatagcgagtttgtgtccctgtactgtgaagtgtgggagatgatcagaag 907
DB 1319 gccctgaagcacatagcgagtttgtgtccctgtactgtgaagtgtgggagatgatcagaag 1378

QY 908 acagcgcccaagctcatccgagaagtgctgatcaagaagctgtctgaagctgtgtgc 967
DB 1379 acagcgcccaagctcatccgagaagtgctgatcaagaagctgtctgaagctgtgtgc 1438
QY 968 tccaattcagaataatccagttgtcttgagaaatctcatgaaagaccctctcttgtgtca 1027
DB 1439 tccaattcagaataatccagttgtcttgagaaatctcatgaaagaccctctcttgtgtca 1498
QY 1028 tcaactgtgcaatccagatggtgtaaagtgtgtccactccacacacaacacagctgt 1087
DB 1499 tcaactgtgcaatccagatggtgtaaagtgtgtccactccacacacaacacagctgt 1558
QY 1088 tccatacctctatgaltctgttgaatacagaanaaacacaacataaagttgtgtgtg 1147
DB 1559 tccatacctctatgaltctgttgaatacagaanaaacacaacataaagttgtgtgtg 1618
QY 1148 caagtgtacttcatcgtgagagctgtgacacactgtgtgaagacctgtgtgaaggtgtgtct 1207
DB 1619 caagtgtacttcatcgtgagagctgtgacacactgtgtgaagacctgtgtgaaggtgtgtct 1678
QY 1208 ccccaagttgatttttgaactgagagatgtgtccagcgttgaatgagatgtctgtga 1267
DB 1679 ccccaagttgatttttgaactgagagatgtgtccagcgttgaatgagatgtctgtga 1738
QY 1268 caactggtctcctctgtanaaatacagctcaaaagttccaagccaagtataaattcttc 1327
DB 1739 caactggtctcctctgtanaaatacagctcaaaagttccaagccaagtataaattcttc 1798
QY 1328 acaagtattccaggaagtataacagcagaagacacagactcaggttttttgaagctaatg 1367
DB 1799 acaagtattccaggaagtataacagcagaagacacagactcaggttttttgaagctaatg 1858
QY 1388 agccagagagaggtgtgaccaaagggaatgttactgtgcagaanaatgtttcatctcgaca 1447
DB 1859 agccagagagaggtgtgaccaaagggaatgttactgtgcagaanaatgtttcatctcgaca 1918
QY 1448 ttacatccacttatagcagctgtccgtgtacacctgtgtggtcatctgtgtgaagcacca 1507
DB 1919 ttacatccacttatagcagctgtccgtgtacacctgtgtggtcatctgtgtgaagcacca 1978
QY 1508 gggcgtgtatgaaagcacctcgcagcagtgatatacaacagcgtcgtccctcctcgaaattcca 1567
DB 1979 gggcgtgtatgaaagcacctcgcagcagtgatatacaacagcgtcgtccctcctcgaaattcca 2038
QY 1568 tcgcgaagaggtcctctctgtgagacaggaatcttgcagaagtgtgaaanaacacacactgagc 1627
DB 2039 tcgcgaagaggtcctctctgtgagacaggaatcttgcagaagtgtgaaanaacacacactgagc 2098
QY 1628 aagaatacttgaaagccataacataatctcctttagagagtgtgcatccattatc 1687
DB 2099 aagaatacttgaaagccataacataatctcctttagagagtgtgcatccattatc 2158
QY 1688 aagagatataatccaatlcagcctcgtgacagaagaatttgaagctttttcaaggtaaaa 1747
DB 2159 aagagatataatccaatlcagcctcgtgacagaagaatttgaagctttttcaaggtaaaa 2218
QY 1748 gcttatalatcaactcagaggaacatacccgattactatttgaacttcttgaacatttg 1807
DB 2219 gcttatalatcaactcagaggaacatacccgattactatttgaacttcttgaacatttg 2278
QY 1808 ccaattgtgcaagtgtctgtgaacttcaataactgtgaactttatggggagctatgctt 1867
DB 2279 ccaattgtgcaagtgtctgtgaacttcaataactgtgaactttatggggagctatgctt 2338
QY 1868 catgvgaaaagcgtgcagagaacacaggtggaatccacatggaagagccccaagaacct 1927
DB 2339 catgvgaaaagcgtgcagagaacacaggtggaatccacatggaagagccccaagaacct 2398
QY 1928 acattcccaagaggtgtatcttcttctcaacttgaagcaggaattcaaggaacttgg 1987
DB 2399 acattcccaagaggtgtatcttcttctcaacttgaagcaggaattcaaggaacttgg 2458
QY 1988 aggtcaactcgggattttagcaagttgaaataagcaagatatacatctgtgggaaaa 2047

|||||
Db 2459 agtcaactccgagattcagcaagtgtgaataagcaagatacatcagatactcgtgga 2518
|||||
Qy 2048 tattcagctcgcacaaagcctcagcgtcgaataaagaatgtgctgtgtgtgaa 2107
|||||
Db 2519 tattcagctcgcacaaagcctcagcgtcgaataaagaatgtgctgtgtgaa 2518
|||||
Qy 2108 gctcagcttctgctcagcagcctgtgaagaacattatctcctcagtggtgaacaaac 2167
|||||
Db 2579 gccctcagttgtctcctcagccctgtgaagaacattatctcctcagtggtgaacaaac 2638
|||||
Qy 2168 cctcaccacatagaagaatgagagcagacatcacatctgttaacaaacctgaaacctgtgaa 2227
|||||
Db 2639 cctcaccacatagaagaatgagagcagacatcacatctgttaacaaacctgaaacctgtgaa 2638
|||||
2228 ttctgacctacagaacatcaagcgtgcgcggtgtgtctgacgtgaagcctgtgttaactga 2287
2699 ttctgacctacagaacatcaagcgtgcgcggtgtgtctgacgtgaagcctgtgttaactga 2738
|||||
Qy 2288 agaaccctacaaagctcataatgagatacaataagaatgaagaagaatgtctataaac 2347
|||||
Db 2759 agaaccctacaaagctcataatgagatacaataagaatgaagaagaatgtctataaac 2818
|||||
Qy 2348 tagctgaagccttgaaaaaaccttgaaagaatgtgtttatttcatttgaccactgtctg 2407
|||||
Db 2819 tagctgaagccttgaaaaaaccttgaaagaatgtgtttatttcatttgaccactgtctg 2878
|||||
Qy 2408 acattgagaggggaatgattacatagtcgaagctctgtcgaagtgaaacctgtgacctg 2467
|||||
Db 2879 acattgagaggggaatgattacatagtcgaagctctgtcgaagtgaaacctgtgacctg 2938
|||||
Qy 2468 aagaacatcaatgattcctcgtcgtctgtcgaatgtcgaatgtcgaatgtcgaatgtcga 2527
|||||
Db 2939 aagaacatcaatgattcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcga 2998
|||||
Qy 2528 atcttcaaatctgtgtcaaaactgtgaacatctctgtatctatcagaagaatcaccgtgaa 2587
|||||
Db 2999 atcttcaaatctgtgtcaaaactgtgaacatctctgtatctatcagaagaatcaccgtgaa 3058
|||||
Qy 2588 atggaatgaagaagctcttcatgatactgcagcagatgaaagtgctgaaagctcagcagc 2647
|||||
Db 3059 atggaatgaagaagctcttcatgatactgcagcagatgaaagtgctgaaagctcagcagc 3118
|||||
Qy 2648 caccgatgctcctcgtgagcgtgtacgtgtcaagcagcagcgtgagcagcgtgttgaacatt 2707
|||||
Db 3119 caccgatgctcctcgtgagcgtgtgtacgtgtcaagcagcagcgtgagcagcgtgttgaacatt 3178
|||||
Qy 2708 tgaagaggtcccaacaactcgtlcaagcttggtgtgaaaaactgtgagactcagatacag 2767
|||||
Db 3179 tgaagaggtcccaacaactcgtlcaagcttggtgtgaaaaactgtgagactcagatacag 3238
|||||
Qy 2768 agatagaatatttggtgcatcttttggaagaacccctctgaaaaactcagcagctgtga 2827
|||||
Db 3239 agatagaatatttggtgcatcttttggaagaacccctctgaaaaactcagcagctgtga 3298
|||||
Qy 2828 attgagcggaacatcggtgagcagtgatgtgctgtcgtcctcagtggtgtatttgaga 2887
|||||
Db 3299 attgagcggaacatcggtgagcagtgatgtgctgtcgtcctcagtggtgtatttgaga 3358
|||||
Qy 2888 atcttaagcaatgtgtttttgacttaglactaaagaattctacgtatccagcat 2947
|||||
Db 3359 atcttaagcaatgtgtttttgacttaglactaaagaattctacgtatccagcat 3418
|||||
Qy 2948 tagtcagaagaacatcagcaaatgttatcagaatgaacttctcgaagaacatcagctgtg 3007
|||||
Db 3419 tagtcagaagaacatcagcaaatgttatcagaatgaacttctcgaagaacatcagctgtg 3478
|||||
Qy 3008 ttgggtggaatctgtatgtatgtatcctcagtggtattatcagctgtg 3052
|||||
Db 3479 ttgggtggaatctgtatgtatgtatcctcagtggtattatcagctgtg 3523
|||||

RESULT 4

AAH9581
ID AAH9581 standard; cDNA; 2950 BP.
XX
AC AAH9581;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:416.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antirheumatic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; vitruide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
KW antiaggregant; haemostatic; vulnary; antileuc; osteopathic; eczema;
KW dermatological; antifungal; antidiabetic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
OS Homo sapiens.
XX
PN WO200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-457603/49.
DR P-PSDB: AAM25640.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS Claim 1; Page 511-512; 1217pp; English.
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; vitruide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianemic; antiaggregant; haemostatic; vulnary;
CC antileuc; osteopathic; dermatological; antiallergic; antidiabetic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 2950 BP; 915 A; 592 C; 628 G; 815 T; 0 other;

Query Match 62.5%; Score 1920.6; DB 22; Length 2950;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1923; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1146 tcgaagtgcattcaatcgcagcctgcagcaactctgtgagacctctctgtgaggtgtgt 1205
    |||
Db 1 tcgaagtgcattcaatcgcagcctgcagcaactctgtgagacctctctgtgaggtgtgt 60
QY 1206 ctcccaagaatttgatttcgaacttcgaggaatgtgtccaggttgaaagagattccctgct 1265
    |||
Db 61 ctcccaagaatttgatttcgaacttcgaggaatgtgtccaggttgaaagagattccctgct 120
QY 1266 gacaaactgtgctctctgttaaatatacaactcaagaagttcaagccaaagtataattctt 1325
    |||
Db 121 gaaactgtgctctctctgttaaatatacaactcaagaagttcaagccaaagtataattctt 180
QY 1326 tcaaaagtcaatccagagatatacagaaggaagaaactcaagcttatttgacgtctca 1385
    |||
Db 181 tcaaaagtcaatccagagatatacagaaggaagaaactcaagcttatttgacgtctca 240
QY 1386 tgaagcagaagaggtgacaaaggggaatgttacttcgcagaataatgtttccattccgga 1445
    |||
QY 241 tgaagcagaagaggtgacaaaggggaatgttacttcgcagaataatgtttccattccgga 300
    |||
Db 301 catatacatcaactatagacagcctgtctcgatcacctgtggtgtcatctgtgaaagccac 360
QY 1506 cagagctgtatagaagcaacctgcagcaagtgtatacaaacagcgtgcctctcggaatttc 1565
    |||
Db 361 cagagctgtatagaagcaacctgcagcaagtgtatacaaacagcgtgcctctcggaatttc 420
QY 1566 catcgcgaagaagggcctctctgtgagagaagaatttcgcaaaagtgtaaaaacccactga 1625
    |||
Db 421 catcgcgaagaagggcctctctgtgagagaagaatttcgcaaaagtgtaaaaacccactga 480
QY 1626 gcaagaagaattctgaaagccataaacaatcctctgttagagttgcgcacatcattata 1685
    |||
Db 481 gcaagaagaattctgaaagccataaacaatcctctgttagagttgcgcacatcattata 540
QY 1686 tcaaaagagatataccaatcagccttcagccaagaatttgaaagcttctcttcaaggttaa 1745
    |||
Db 541 tcaaaagagatataccaatcagccttcagccaagaatttgaaagcttctcttcaaggttaa 600
QY 1746 aagcttatatacaactcaggggaacatcccccgttaacttatgtacttcttgaacattt 1805
    |||
Db 601 aagcttatatacaactcaggggaacatcccccgttaacttatgtacttcttgaacattt 660
QY 1806 gcccaattgtgcagagtctctgtgacttcaactgcgaatttlatggtgggagactatgac 1865
    |||
    661 gcccaattgtgcagagtctctgtgacttcaactgcgaatttlatggtgggagactatgac 720
    |||
    1866 ttcataaggaaagagctgcagaaagacaaggttgaaatccacatgaaagagccccagaac 1925
    |||
    721 ttcataaggaaagagctgcagaaagacaaggttgaaatccacatgaaagagccccagaac 780
    |||
QY 1926 ctacattcccaagagggcgtatcttcttcaacttgaaagcaagaatttcagagctct 1985
    |||
Db 781 ctacattcccaagagggcgtatcttcttcaacttgaaagcaagaatttcagagctct 840
    |||
QY 1986 ggaaggtcacacaccggagatttcagcaagttgaaataagcaagatatcacatattcgggaa 2045
    |||
Db 841 ggaaggtcacacaccggagatttcagcaagttgaaataagcaagatatcacatattcgggaa 900
    |||
QY 2046 aatatcagctctgcacaaagcctcaggtgcacaataaagagaatgtcgtgtgtgagctgg 2105
    |||
Db 901 aatatcagctctgcacaaagcctcaggtgcacaataaagagaatgtcgtgtgtgagctgg 960
    |||
QY 2106 aagccctcaattgtgtcctcagcaccgttagaacatttctctcatagtgtgaagccag 2165
    |||
Db 961 aagccctcaattgtgtcctcagcaccgttagaacatttctctcatagtgtgaagccag 1020
    |||
```

```
QY 2166 tccctcaccatagaagatgagagcacaatcacatctgttaacaacaacctgaaaccttgag 2225
    |||
Db 1021 tccctcaccatagaagatgagagcacaatcacatctgttaacaacaacctgaaaccttgag 1080
    |||
QY 2226 tattcatgacctcagaatatacaacggtctgcgggtgtgtctgtactgtacaagcttgggttaact 2285
    |||
Db 1081 tattcatgacctcagaatatacaacggtctgcgggtgtgtctgtactgtacaagcttgggttaact 1140
    |||
QY 2286 gaaagaaccttacaaaggtctcaatgaaatacaataaagaatgaagaagaatgtcttaaa 2345
    |||
Db 1141 gaaagaaccttacaaaggtctcaatgaaatacaataaagaatgaagaagaatgtcttaaa 1200
    |||
QY 2346 actagctgaagccttgaaaaaccttgaaagaagatgtgttatttcatcttgagccactgttc 2405
    |||
Db 1201 actagctgaagccttgaaaaaccttgaaagaagatgtgttatttcatcttgagccactgttc 1260
    |||
QY 2406 tgcacattggaagaggaatgtgattacataagtcacaaagctctgttcaagtgtaacctgtgaact 2465
    |||
Db 1261 tgcacattggaagaggaatgtgattacataagtcacaaagctctgttcaagtgtaacctgtgaact 1320
    |||
QY 2466 tgaagaaatccaattagctcctcgctgtgtgtgcaaaatgcaggtgaaaaatccttagctca 2525
    |||
Db 1321 tgaagaaatccaattagctcctcgctgtgtgtgcaaaatgcaggtgaaaaatccttagctca 1380
    |||
QY 2526 gaactctcaacaattgtgtcaaaactgagcaatctctgtatttatacgaaaatataccttgaaaa 2585
    |||
Db 1381 gaactctcaacaattgtgtcaaaactgagcaatctctgtatttatacgaaaatataccttgaaaa 1440
    |||
QY 2586 agatgtgaatataagctcttcaatgaactgtatcgacaggaatgaacgtgtctagaagactcac 2645
    |||
Db 1441 agatgtgaatataagctcttcaatgaactgtatcgacaggaatgaacgtgtctagaagactcac 1500
    |||
QY 2646 cgcacatgactgtcctctggtggcgtgtgacgtgtgcaaggcaagccttgaaagcctgtgttgaaac 2705
    |||
Db 1501 cgcacatgactgtcctctggtggcgtgtgacgtgtgcaaggcaagccttgaaagcctgtgttgaaac 1560
    |||
QY 2706 ttgtgagaagagtcaccaaacactgcataagcttgggttgaaaaacttgagactcacaataac 2765
    |||
Db 1561 ttgtgagaagagagtcaccaaacactgcataagcttgggttgaaaaacttgagactcacaataac 1620
    |||
QY 2766 agagattagaatttttagtgacatttttttgaaagaacacctctgaaaaactctcagaagtt 2825
    |||
Db 1621 agagattagaatttttagtgacatttttttgaaagaacacctctgaaaaactctcagaagtt 1680
    |||
QY 2826 gaatttgcggggaatcgtgtgagcagtgaaatgagtggtgcttcataaggttatatta 2885
    |||
Db 1681 gaatttgcggggaatcgtgtgagcagtgaaatgagtggtgcttcataaggttatatta 1740
    |||
QY 2886 gaactctaagcaattagtggttttttgactttagtactaaagaaatttctactatccagc 2945
    |||
    1741 gaactctaagcaattagtggttttttgactttagtactaaagaaatttctactatccagc 1800
    |||
QY 2946 attagtcagaanaacttagccaagtgttataccaagttaactttctcagaagaactagct 3005
    |||
    1801 attagtcagaanaacttagccaagtgttataccaagttaactttctcagaagaactagct 1860
    |||
QY 3006 tgttgggttgcaatttgatgataatgataatcgaatgttatatacaggtgtctttaaactagt 3065
    |||
    1861 tgttgggttgcaatttgatgataatgataatcgaatgttatatacaggtgtctttaaactagt 1920
    |||
    3066 aactgct 3072
    |||
    1921 aactgct 1927
    |||
```

```
RESULT 5
AAH34171/c
ID AAH34171 standard; cDNA; 2725 BP.
XX
XX AAH34171;
AC
XX
DT 03-SEP-2001 (first entry)
XX
```


CC Include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 421 BP; 126 A; 69 C; 77 G; 149 T; 0 other;

Query Match

7.9%; Score 242; DB 22; Length 421;

Best Local Similarity 100.0%; Pred. No. 3.8e-61;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2781 aggtgcatcttttggaaagaaccctcgaaaactccagcaggttgattggcgggaa 2840
|||||
180 aggtgcatcttttggaaagaaccctcgaaaactccagcaggttgattggcgggaa 239
|||||
QY 2841 tcgtgtgagcagtgatgagtgctgcctcatgggtgattgagaactcagaact 2900
|||||
DB 240 tcgtgtgagcagtgatgagtgctgcctcatgggtgattgagaactcagaact 299
|||||
QY 2901 agtgtttttgacttagtactaagaattcttaccctgatacagcattagtcagaaact 2960
|||||
DB 300 agtgtttttgacttagtactaagaattcttaccctgatacagcattagtcagaaact 359
|||||
QY 2961 tagccaagtgatataccaacttcttcgaagaagctagctgttggttgcaatt 3020
|||||
DB 360 tagccaagtgatataccaacttcttcgaagaagctagctgttggttgcaatt 419
|||||
QY 3021 tg 3022
||
DB 420 tg 421

RESULT 9

AA123590
ID AA123590 standard; DNA; 220 BP.

AC AA123590;

DT 12-OCT-2001 (first entry)

DE Probe #13523 for gene expression analysis in human cervical cell sample.

KW Probe; human; microarray; gene expression; cervical epithelial cell;
cervical cancer; ss.

XX Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 13523; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
(SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match

7.2%; Score 220; DB 22; Length 220;

Best Local Similarity 100.0%; Pred. No. 8.1e-55;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 ttggcggggaatcgtgtgagcagtgatgagtgctgcctcatgggtgatttgagaat 2889
|||||
DB 1 ttggcggggaatcgtgtgagcagtgatgagtgctgcctcatgggtgatttgagaat 60
|||||
QY 2890 cttaagcaattagtggtttttgacttagtactaagaattcttaccctgatacagcata 2949
|||||
DB 61 cttaagcaattagtggtttttgacttagtactaagaattcttaccctgatacagcata 120
|||||
QY 2950 gtccagaaacttagcccaagtgatataccaagttacttctgcaagaagctagctgtt 3009
|||||
DB 121 gtccagaaacttagcccaagtgatataccaagttacttctgcaagaagctagctgtt 180
|||||
QY 3010 gggtggcaatttgatgatgatatactccaagtgattatacg 3049
|||||
DB 181 gggtggcaatttgatgatgatatactccaagtgattatacg 220
|||||

RESULT 10

AA148904
ID AA148904 standard; DNA; 220 BP.

AC AA148904;

DT 17-OCT-2001 (first entry)

DE Probe #17590 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;
genetic disorder; ss.

XX Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-48897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 17590; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match 7.28; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 8.1e-55;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 ttggcgggaaatcgctgagcagtgatgagtgcttccttcacgggtgtatttgaagaat 2889
Db 1 ttggcgggaaatcgctgagcagtgatgagtgcttccttcacgggtgtatttgaagaat 60

QY 2890 ctaagcaattagtgcttttttgacttagtactaaagaatttctactgacccagcatla 2949
61 ctaagcaattagtgcttttttgacttagtactaaagaatttctactgacccagcatla 120

QY 2950 gtcaagaaacttagccaagtgtatccaagttaactttctgcagaagaagctagctgtt 3009
121 gtcaagaaacttagccaagtgtatccaagttaactttctgcagaagaagctagctgtt 180

QY 3010 ggggtggcaattgtagtgatgatactcaagtgatttaccag 3049
Db 181 ggggtggcaattgtagtgatgatactcaagtgatttaccag 220

RESULT 11
AA109206
ID AA109206 standard; DNA: 220 BP.
XX
AC AA109206;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #9197 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast; disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN MO200157270-A2.
XX
XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID NO 9197; 322bp; English.

XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 220 BP; 58 A; 32 C; 54 G; 76 T; 0 other;

Query Match 7.28; Score 220; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 8.1e-55;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 ttggcgggaaatcgctgagcagtgatgagtgcttccttcacgggtgtatttgaagaat 2889
Db 1 ttggcgggaaatcgctgagcagtgatgagtgcttccttcacgggtgtatttgaagaat 60

QY 2890 ctaagcaattagtgcttttttgacttagtactaaagaatttctactgacccagcatla 2949
61 ctaagcaattagtgcttttttgacttagtactaaagaatttctactgacccagcatla 120

QY 2950 gtcaagaaacttagccaagtgtatccaagttaactttctgcagaagaagctagctgtt 3009
121 gtcaagaaacttagccaagtgtatccaagttaactttctgcagaagaagctagctgtt 180

QY 3010 ggggtggcaattgtagtgatgatactcaagtgatttaccag 3049
Db 181 ggggtggcaattgtagtgatgatactcaagtgatttaccag 220

RESULT 12
AAF58252
ID AAF58252 standard; DNA: 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
XX 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX

[illegible][illegible]

RESULT	14
AAFS8257	
ID	AAFS8257 standard; DNA; 936 BP.
XX	
AC	AAFS8257;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1954.
XX	
KM	Electron-transfer group; ETW; mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.
XX	
PN	W0200107665-A2.
XX	
PD	01-FEB-2001.
XX	
XX	
PF	26-JUL-2000; 2000MO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.

QY 2640 gctcacccgactgactgcctcgggctgacgtgcaaggcagcctgaacgctgt 2699
Db 377 www.wwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwww 436

QY 2700 gaacattggaggaaggtcccaacactgtcaagcttgggttgaaacactggagactcac 2759
Db 437 www.ww 496

QY 2760 agatacagagattagaatttagtgcatttttggaaagacccctgaaaaactcca 2819
Db 497 www.ww 556

QY 2820 gcagttgaattbgcgggaatcgctgagcagtgatgctgccttcaatgggtgt 2879
Db 557 www.ww 616

QY 2880 attggaactctaagaaattggttlttgaacttgactaagaattctactoga 2939
Db 617 www.ww 676

QY 2940 tccaagcatgctagaanaacttagccaagtgttacccaagttaactttctgcgaagaagc 2999
Db 677 www.ww 736

DR	WPI; 2001-159728/16.
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
XX	
PS	Example 6; Page 127; 159pp; English.
XX	
CC	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
QQ	Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

	RESULT	15
	AAF58259	
ID	AAF58259	standard; DNA; 936 BP.
XX		
XX	AAF58259;	
AC		
DT	24-APR-2001	(first entry)
XX		
DE	Oligonucleotide D2004.	
XX		
KW	Electron-transfer group; ETM; mismatch; genotyping;	
KM	gene expression; ss.	
XX		
SS	Synthetic.	

Query Match 3.0%; Score 91.8; DB 22; Length 936;

W0200107665-A2.

```
PD 01-FEB-2001.
XX
XX 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
XX MPI: 2001-159728/16.
XX
DR Nucleic acids containing electron-transfer group, useful as labels in
PI hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
XX Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 3.0%; Score 91.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.6e-16;
Matches 6; Conservative 452; Mismatches 309; Indels 0; Gaps 0;

QY 2280 taacttgaagaaccttacaagaatcataatgataacataaagaatgaagaagatgc 2339
DB 17 www. .... 76
QY 2340 tataaactagctgaagccctgaaaaacctgaagaagatgcttatttcattgaccca 2399
DB 77 www. .... 136
QY 2400 ctgtctgacattggaagggaatgatataatgataagctctgtcaagtgaaccctg 2459
DB 137 www. .... 196
QY 2460 tgaacttgaagaatcctaattagctcctgtctgtctgcaaatgcaatgaaatcct 2519
DB 197 www. .... 256
QY 2520 agctcagaatcttcacaattgttcaaatgcaatcctctgtattatcagaatattact 2579
DB 257 www. .... 316
QY 2580 ggaaaaagatggaatgaagctctcattgaaactgatacgacagatgaacgtgtagaaca 2639
DB 317 www. .... 376
QY 2640 gctcaccgaatgctgcgcctgtggctgtgacgtgcaaggcaagcctgagcagcctgtt 2699
DB 377 www. .... 436
QY 2700 gaacatttggagggtcccaacaactcgtcaagcttggtgtaaaaactgagactcac 2759
DB 437 www. .... 496
QY 2760 agatacagaatagatatttagtgcatttttggaaagaacctctgaaactcca 2819
DB 497 www. .... 556
QY 2820 gcagttgaatttgcgcggaaatcgtgtgagcagtgatgagtgctgccttcattggtgt 2879
DB 557 www. .... 616
```

```
QY 2880 atttgaagaatctaaagcaattagtgctttttagactttagtactaaagaatttctacciga 2939
DB 617 www. .... 676
QY 2940 tccagcattagtcagaaagaacttagccaagtgtatccaaagttaactttctgcaagaagc 2999
DB 677 www. .... 736
QY 3000 taggcttgttggtggtgcaatttgatgatgatctcagtgcttata 3046
DB 737 www. .... 783
```

Search completed: March 25, 2002, 11:45:30
Job time: 3314 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2002, 11:37:02 ; Search time 2764.94 Seconds

(without alignments)
11939.149 Million cell updates/sec

Title: US-09-697-089-3

Sequence: 3072
1 atgaattcattcaagagacaa.....ctttaactagtaactgct 3072

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estopl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: qb_estli:*
11: qb_est2:*
12: qb_hic:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	551	17.9	552	13	AO309404 CITBI-EI-
C 2	466.4	15.2	480	10	AV719179
C 3	407.8	13.3	602	13	AZ360053
C 4	347	11.3	371	10	AV656315
C 5	342	11.1	364	10	AI263294
C 6	285.8	9.3	476	11	BG210375
C 7	282.8	9.2	470	13	AO624020
C 8	223.8	7.3	404	13	AO889169
C 9	223	7.3	261	10	AW337918
C 10	170	5.5	553	13	AO320928
C 11	157.2	5.1	219	13	AO283886
C 12	153.8	5.0	251	11	BF903662

C 13	146	4.8	630	13	AO112439	AO112439 CIT-HSP-2
C 14	139.8	4.6	509	11	BF207840	BF207840 601862546
C 15	136	4.4	540	10	AW418826	AW418826 ha21e11.x
C 16	106	3.5	728	13	AZ720059	AZ720059 RCTI-24-8
C 17	98.8	3.2	412	11	H25984	H25984 Y156g07.r11
C 18	93.4	3.0	499	10	AI023795	AI023795 ox08d03.x
C 19	66.2	2.2	429	13	AZ484615	AZ484615 IM0311P08
C 20	55	1.8	635	13	B58691	B58691 CIT-HSP-201
C 21	49.2	1.6	668	13	AZ762115	AZ762115 IM0556N15
C 22	46.8	1.5	625	13	AZ614134	AZ614134 IM0442N17
C 23	44.8	1.5	546	13	AZ362463	AZ362463 IM0107N03
C 24	42.2	1.4	1023	10	BE306291	BE306291 601103781
C 25	40.8	1.3	835	10	BE643259	BE643259 Cr12_8.FO
C 26	40.6	1.3	361	10	AW504276	AW504276 UT-HF-BNO
C 27	40.2	1.3	783	11	BG563851	BG563851 602584655
C 28	40	1.3	424	13	AO638259	AO638259 927P1-17A
C 29	40	1.3	518	10	AA669206	AA669206 ad93b07.s
C 30	39.8	1.3	939	11	BE887277	BE887277 601508592
C 31	39.6	1.3	467	10	AW988289	AW988289 u05f04.y
C 32	39.4	1.3	363	10	BE544063	BE544063 601069826
C 33	39.4	1.3	523	11	BF259332	BF259332 HVSMEF001
C 34	39.4	1.3	549	11	BG109381	BG109381 602280435
C 35	39.4	1.3	573	11	BF257885	BF257885 HVSMEF001
C 36	39.4	1.3	611	11	BF253881	BF253881 HVSMEF000
C 37	39.4	1.3	754	11	BF258252	BF258252 HVSMEF001
C 38	39.4	1.3	854	11	BE881131	BE881131 601492001
C 39	39.2	1.3	759	10	AI729528	AI729528 BNGH1135
C 40	39.2	1.3	806	13	AO481970	AO481970 GNG1662.m
C 41	39.2	1.3	1101	13	CNS016LW	AL106910 Drosoph11
C 42	39	1.3	1201	13	CNS007JU	AL067358 Drosoph11
C 43	38.6	1.3	574	10	AW658668	AW658668 95275.MAR
C 44	38.4	1.2	425	10	AI358206	AI358206 gV96907.x
C 45	38.4	1.2	640	11	BI391271	BI391271 ppp1n.pk0

ALIGNMENTS

RESULT 1
LOCUS AO309404 552 bp DNA
DEFINITION CITBI-EI-2528J13.TF CITBI-EI Homo sapiens genomic clone 2528J13, DNA sequence.
ACCESSION AO309404.1 GI:4041438
VERSION AO309404.1
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other-GSSs: CITBI-EI-2528J13.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (Info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
Location/Qualifiers
1..552
/organism="Homo sapiens"

FEATURES
source

```

/db_xref="taxon:9606"
/clone="2528J13"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT      144 a      121 c      130 g      157 t
ORIGIN

Query Match      17.9%; Score 551; DB 13; Length 552;
Best Local Similarity 100.0%; Pred. No. 1.2e-135;
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1427 aaatggttcacatttcgacattacacactatagcagctgcggtacacctgtg 1486
Db 552 AAATGTTTCATTTCGCGACATTACATCCATTATAGCAGCTGCTCGGTACCTGTG 493

Oy 1487 ggtcacctgtgaaagccaccagggctgtatgaagacctcgacagcagtgtalcaacag 1546
Db 492 GGTCACTGTGTGAAGCCACACAGGGGCTGTATGAAGCACCTCGCACAGTGTATCACACAG 433

Oy 1547 gctgctctcgtgacttttcacgcacaaaggcctctctgagacaggaatcttgcacaa 1606
Db 432 GCTGCTCTCTCGACTTCATCGCCAGAGAGCCTCTCTGAGACAGGATCTTGCCAAA 373

Oy 1607 gttgtgaaaaaacacacacgacgaagaatctgaagaacataacatcctctgtag 1666
Db 372 GGTGTAAGAAACACACACGACGACGAAGAAATCTGAAGCCATTAACATCTCTTTGTAG 313

Oy 1667 agtgtgacatcattatatacaagagatgacatccaatcagccctgagccaagaattg 1726
Db 312 AGTGTGGCATTCATTATATCAAGAGATACATCCAAATCAGCCCTGAGCCAAAGATTGG 253

Oy 1727 aagcttctctcaaggttaaaagcttatatacaactcaggaacatccccgattactat 1786
Db 252 AAGCTTCTCTTAAGGTAAAGCTTATATATCAACTCAGGAAACATCCCGATTACTTAT 193

Oy 1787 ttgacttctcttgaaacatttgcacaaattgtgcagtgctctgacatcattaaactgact 1846
Db 192 TTGACTTCTTTGAACATTGTGCCAATTGTGCAAGTGCCTGTGACTCATTAACGTGACT 133

Oy 1847 ttatgtggggagactatgtctcatgtggaaaggtcgacagaacacaggttgaaactaca 1906
Db 132 TTTATGGGGGACCTATGTGCTTCATGGGAAAGGCTGCACAAACACAGGTGATCCACA 73

Oy 1907 tggaaagagcccccagaacactatctccagaggggttatcttgtcttcaactgaa 1966
Db 72 TGGAAAGAGGCCCCAGAAACCTACATTCCAGCAGAGGCTGTATCTTGTTCATCTGGA 13

1967 agcaggaattc 1977
Db 12 AGCAGGAATTC 2

RESULT 2
AV719179 480 bp mRNA EST 16-OCT-2000
LOCUS AV719179 GLC Homo sapiens cDNA clone GLOCDA10 5', mRNA sequence.
DEFINITION AV719179
ACCESSION AV719179
VERSION AV719179.1 GI:10816331
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 480)
AUTHORS Qian,B., Wu,T., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GLC clones

```

```

JOURNAL Unpublished (2000)
COMMENT Contact: Zenguan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Source 1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCEDA10"
/clone_lib="GLC"
/tissue_type="Corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      138 a      120 c      85 g      137 t
ORIGIN

Query Match      15.2%; Score 466.4; DB 10; Length 480;
Best Local Similarity 99.8%; Pred. No. 4e-113;
Matches 467; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2605 catgaactatgcagacaggttgaaagctgctagaaacactcacccgactgctgacccgtg 2664
Db 480 CATGAACCTGATGCACAGATGAAAGCTGTAGAACACTCACCCACTGATGCTGCTGG 421

Oy 2665 ggtctgtaagctgtaagagcagcctgtaagacccctgttgaacatttgaggaagtcacacaa 2724
Db 420 GGCTGTACCTGCAAGACGACGCTGAGCAGCAGCTGTGAAACATTTGGAGAGGTCACAAA 361

Oy 2725 ctctgtaagcttggtgttgtaaaacttgagagctcacagatacagagattgaatttagt 2784
Db 360 CTCTGTAACCTTGGGTTGAAAACTGAGACCTCACAGATACAGAGATTGAAATTTAGGT 301

Oy 2785 gcaatttcttgaaagaacccctctgaaaaacttcagcagattgaattgctcggaagactgt 2844
Db 300 GCATTTTGGAAAGAACCCCTCTGAAAACTTCACACAGTTGAATTTGGCGGAATTCGT 241

Oy 2845 gtgagcagtgatgagatggttgcttcacatggtgtgtaattgagaacttgaactagtg 2904
Db 240 GTGACGAGTGATGATGGCTTGCTTCATGGGTGTATTGAGAACTTTAAGCAATTAAGTG 181

Oy 2905 tttttgactttagtactaagaattctaacctgataccagcatagtcagaaacttagc 2964
Db 180 TTTTGTGACTTTAGTACTAAAGAAATTTCTACCTGATCCAGCATTAAGTACGAAAACTTAGC 121

Oy 2965 caagtgtatccaaagtaactttctgcaagaagcctgctgttggtggtgcaatttgat 3024
Db 120 CAAGTGTATTCAGATTAACTTTCTGCAAGAGAGCTAGAGCTGTGTGTGGCGCAATTGAT 61

Oy 3025 gatgatgactcagtgatgtaatacaggtgctttaaactgtaactgct 3072
Db 60 GATGATGATCTCAGTGTATTATACAGAGAGCTTTAAACTGATTAACCTCT 13

RESULT 3
AZ360053 602 bp DNA GSS 02-OCT-2000
LOCUS AZ360053/c 1M0103H11F Mouse 10kb plasmid UNGC1M library Mus musculus genomic
DEFINITION clone UNGC1M0103H11 F, DNA sequence.
ACCESSION AZ360053
VERSION AZ360053.1 GI:10473753
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

```


QY 1933 cccagcaggcgtctatcttcttctcaactggaagcaagaaattcaggactctgagtc 1992
|||||
Db 61 cccagcaggcgtctatcttcttctcaactggaagcaagaaattcaggactctgagtc 120
QY 1993 acaactcggagattcagcaagattgaaataagcaagatacatactctgggaaataatc 2052
|||||
Db 121 acacttcgggatttcagcaagattgaaataagcaagatacatactctgggaaataatc 180
QY 2053 agctctgcacaagcctcagcagctgcaataaagagatgctggtgtgtgagcctc 2112
|||||
Db 181 agctctgcacaagcctcagcagctgcaataaagagatgctggtgtgtgagcctc 240
QY 2113 agttgttctcagcagcctctgtaagaaatttctctcatggtgtggaagcagtcctc 2172
|||||
Db 241 agttgttctcagcagcctctgtaagaaatttctctcatggtgtggaagcagtcctc 300
QY 2173 accatagaagatgagagggcacatcatctgttaacaacacctgaagta 2227
|||||
Db 301 accatagaagatgagagggcacatcatctgttaacaacacctgaagta 355

RESULT 5

LOCUS AI263294 364 bp mRNA EST 03-FEB-1999
DEFINITION qx57b01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2005417 3',
mRNA sequence.
ACCESSION AI263294
VERSION AI263294.1 GI:3871497
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 364)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-ri@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 2146 std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 364.
Location/Qualifiers
1..364

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Pan1"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: PCWV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 117 a 84 c 55 g 108 t
ORIGIN

Query Match 11.1%; Score 342; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 5.1e-80;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2731 aaagcttggttgaagaaacgagacacacagatagagattgaattgaattc 2790
|||||
Db 364 aaagcttggttgaagaaacgagacacacagatagagattgaattgaattc 305

QY 2791 ttggaagaagaccctcbyaaaacctccagcagltgaattggcggaactcgtgagc 2850
|||||
Db 304 ttggaagaagaccctcbyaaaacctccagcagltgaattggcggaactcgtgagc 245
QY 2851 agtgaatgagcttgccttcacatggtgtatlttgagaattcttaagcaattagtttt 2910
|||||
Db 244 agtgaatgagcttgccttcacatggtgtatlttgagaattcttaagcaattagtttt 185
QY 2911 gacttgaactaaagattcctcaccatccagcagatgaacgaacaaagtg 2970
|||||
Db 184 gacttgaactaaagattcctcaccatccagcagatgaacgaacaaagtg 125
QY 2971 ttatccaagttaacattcttcgcaagaagcaggctgtgtgtggtgcaattgagatg 3030
|||||
Db 124 ttatccaagttaacattcttcgcaagaagcaggctgtgtgtggtgcaattgagatg 65
QY 3031 gatctcagtgattatcaggtgctttaaactgtaactgct 3072
|||||
Db 64 gatctcagtgattatcaggtgctttaaactgtaactgct 23

RESULT 6

LOCUS BG210375 476 bp mRNA EST 21-APR-2001
DEFINITION RSJ29913 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG210375
VERSION BG210375.1 GI:13732062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 476)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramchandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velooso,N., Hess,J., Colchen,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 360.
Location/Qualifiers
1..476

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT 120 a 99 c 107 g 149 t
ORIGIN

Query Match 9.3%; Score 285.8; DB 11; Length 476;
Best Local Similarity 97.6%; Pred. No. 4.9e-65;
Matches 290; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2776 atttgatgcatcttttggaaagacccctgaaaaactcagcagtgatggcg 2835
|||||
Db 145 actatagtgatcttcttgaaagacccctgaaaaactcagcagtgatggcg 204
QY 2836 ggaatcgtgtgagcagtgatgagcttgcctcatggtgatttggaatcttaag 2895

|||||
Db 205 GGAATTCGTCAGCAGATGATGATGGCTTGCTGGTGCGTGTATTGCAATCTTAAG 264
Oy 2896 caattagagtttttgcagcttctgacttaagaattctaccgcatccgcatatgcga 2955
|||||
Db 265 CAATTAGTGTCTTTTGTGACTTACTTAAGAAATTTCTACATGCATCCACATTAATGTCAGA 324
Oy 2956 aaacttaacaaagtgttaccagatcattcttcgcaagaagcgaagctgtgtggtg 3015
|||||
Db 325 AAACCTAGCAGATGATGATCTCAGTCTATTACAGGCTTTTAACTAGTACTGCT 384
Oy 3016 caattgagatgagatcagctcagtggttattacaggtgtctttaactagtaactgt 3072
|||||
Db 385 CAATTGATGATGATGATCTCAGTCTATTACAGGCTTTTAACTAGTACTGCT 441

RESULT 7
A0624020

LOCUS A0624020 470 bp DNA GSS 16-JUN-1999
DEFINITION HS_5378_B2_C12.SP6E RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION A0624020
VERSION A0624020.1 GI:5086412
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 470)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
JOURNAL scanning the human genome
MEDLINE Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENT Contact: Mahairas GC, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end web Server:
http://www.htsc.washington.edu
Plate: 954 row: F column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 470.

FEATURES

source
1. 470
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=954 Col=24 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRII. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 142 a 101 c 107 g 117 t 3 others
ORIGIN

Query Match 9 2% Score 282.8; DB 13; Length 470;
Best Local Similarity 84.1%; Pred. No. 3.1e-64;
Matches 317; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Oy 1972 gaattcagactctgagagtcacactccggaatttcagcaagttgaataagaatalc 2031
|||||
Db 2 GATATCATGACCTCTGGAGGTCTACACTCCGGGATTTTCAGCAACTGATAGCAAGATATC 61
Oy 2032 acatactcgggggaaataatctcagctctgcacaaagcctcgaagctcgaataagagatgt 2091
|||||
Db 62 AGATATCTGGGGGAAAATATTCAGTCTCTGCCACAACTCAGGCTGCAATAAAGAGATGT 121
Oy 2092 gctggtgtggtcgggaagcctcagttgtgtctcgaagcctcgaataaactatctctc 2151
|||||
Db 122 GCTGTGTGTGGCTGGAAAGCCTCAGTTGTGCTCAGCAGCCTGTAAAGAACATTATTTCTTC 181
Oy 2152 atcgtcgaagcagcagccctccacatagaagaatgaagagacacatctcgtlaacaac 2211
|||||
Db 182 ATGTGGAAAGCCAGTCGCTCAGTCCATGAGATGAGAGGACATCATCTGTAAATATTC 241
Oy 2212 ctgaagaacctgagatcattcatcagctcagaaatcagaagctcgggtgtgtcactgaac 2271
|||||
Db 242 CTGAAGAACCTTGAGTATTCATGACCTACAGATCAACGCTCCGGGTATTTGTAATATTC 301
Oy 2272 agcttggtgacttgagaagacctcaagcctcatalgtatgataagaatgaatgaa 2331
|||||
Db 302 AGTGCTTTGTGCTTGTTCATTACATTAATAATACAGTGTATTAACCTTGAAAAAGAGA 361
Oy 2332 gaagatgctataaact 2348
|||||
Db 362 GAGGAGACTTTATTCCT 378

RESULT 8

LOCUS A0889169 404 bp DNA GSS 10-NOV-1999
DEFINITION HS_2161_B1_A01_77C CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2161 Col=1 Row=B, DNA sequence.

ACCESSION A0889169
VERSION A0889169.1 GI:6345359
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 404)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
JOURNAL scanning the human genome
MEDLINE Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENT Contact: Mahairas GC, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end web Server: http://www.htsc.washington.edu
Plate: 2161 row: B column: 1
Seq primer: T7
Class: BAC ends
High quality sequence stop: 404.

FEATURES

source

1. 404
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2161 Col=1 Row=B"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 119 a 93 c 70 g 122 t
ORIGIN

Query Match 7.3%; Score 223.8; DB 13; Length 404;
Best Local Similarity 91.5%; Pred. No. 1.5e-48;
Matches 248; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

QY 2 tgaattcataaaggacataagccgagcccttattcaagaatlyggaactgtttaa 61
DB 350 TGAATTCAATAGACACAGATAGCCGAG-CCTTATTTCAAAAGATGGATGCTTTATAG 292
QY 62 agcaaatcacagatgaccttttgatggaatgttcggaatcggaagaagtaacta 121
DB 291 AGCTACTCACAGATGACCTTTTGTATGAATGTTATGAATCGAAGAAAGTAAACATCA 232
QY 122 ttctgctgagaaggctgagcagatgctgtagaaggatcatcatgatttga 181
DB 231 TTTTCTGGAGAAAGGTGAGACAGATGCTGTAGAGGATCATTCACATGATTTGAAG 172
QY 182 agggctcagaagctgtgaacctttcttaaaccttaaggagtggaactatctat 241
DB 171 AGGTTTCAGAGTCTCTTACCTCTATGTTAAATCCCTTAAGAGACTGAACATCTCTAT 112
QY 242 ttcaaggactgaatgagcaaatgttttca 272
DB 111 TTCAGGACTTGAAGGACAAAGTAAATGATATCA 81

RESULT 9
AM337918/c 261 bp mRNA EST 31-JAN-2000
LOCUS he12h11.x1 NCI-CGAP-CHL1 Homo sapiens cDNA IMAGE:2918853 3',
DEFINITION mRNA sequence.
ACCESSION AM337918 GI:6834544
VERSION AM337918
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 261)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabrs@email.nih.gov
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdnp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 201.
Location/Qualifiers
1. 261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2918853"
/clone_1ib="NCI-CGAP-CHL1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1;
Salt: Site: 2; NotI; Cloned unidirectionally. Primer:
Oligo dT. Library constructed by Life Technologies."

BASE COUNT 86 a 54 c 38 g 83 t

ORIGIN

Query Match 7.3%; Score 223; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2850 cagatgatgagtgcttgccttcattgggtgattttgaatccttaagaattaggtttt 2909
DB 261 CAGTATGATGATGGCTTCCCTTCATAGGCTGTATTTGAAGAACTTAAAGCAATAGGTTTTT 202
QY 2910 tgaattagtaactaaagaatttcaactgattccagatcagatcagaactagcaagt 2969
DB 201 TGAATTAGTACTTAAAGAAATTTTACTGATCCAGCATTTAGTGAAGAACTTACCAAGT 142
QY 2970 gtatccaaagttaactttcttcgaagaagcagatgctgttggtlygcaattgatga 3029
DB 141 GTATTCAGATTAACCTTTTCTGCAAGAGCTAGGCTGTGGGCGCAATTGATGATGA 82
QY 3030 tgatctagtgattatcaggtgcttttaactagtaagct 3072
DB 81 TGATCTCAGTGTATTAACAGTCTCTTTAAACTAGTAAGTCT 39

RESULT 10
AQ320928/c 553 bp DNA GSS 06-MAY-1999
LOCUS RPI11-93C9.TV RPI1-11 Homo sapiens genomic clone RPI1-11-93C9, DNA
DEFINITION sequence.
ACCESSION AQ320928
VERSION AQ320928
KEYWORDS AQ320928.1 GI:4053662
SOURCE GSS.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 553)
Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
Berry K., Granger D., Suh E., White C., de Jong P. and Venter J.C.
Use of Human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other GSSs: RPI11-93C9.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetfgr.org

clones are derived from the human BAC library RPI1-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buhalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buhalo.edu/ordering>) or from
Research Genetics (inforesgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: 77
Class: BAC ends.
Location/Qualifiers
1. 553
/organism="Homo sapiens"
/db_xref="GDB:753384"
/db_xref="taxon:9606"
/clone="RPI1-11-93C9"
/clone_1ib="RPI1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: PBACE3.6; Site: 1; EcoRI; Site: 2; EcoRI;
RPI11 Human Male BAC Library"

BASE COUNT 170 a 107 c 114 g 162 t

ORIGIN

Query Match 5.5%; Score 170; DB 13; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.4e-34;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2615 tcgacagatgaacgtgcttagaacagctcaccgcacgtatgctccctggtggtgacg 2674
 |||||||
 Db 258 TCGACAGATGAACCTGCTAGAAAGCTCACCGCACTGATGCTGCCCTGGGGCTGTGACG 199
 |||||||
 Qy 2675 tgcagaagcagctgagcagcctgtgtgaaacatttgagagaggtcccaacacgtcgaagc 2734
 |||||||
 Db 198 TGCAGAGCAGCTGAGCAGCTGTGAAACATTGTGAGAGAGTCCACAAACGTCGAAGC 139
 |||||||
 Qy 2735 ttgggttcaaaaactgagagactcacagatataagattagattagct 2784
 |||||||
 Db 138 TTGGCTTAAAAAAGCTGAGACTCAGACATACAGATTAAGATTAGT 89
 |||||||

RESULT 11
 AQ28386 219 bp DNA GSS 27-APR-1999
 LOCUS RPII11-78E13.TV RPI1-11 Homo sapiens genomic clone RPI1-11-78E13,
 DEFINITION DNA sequence.
 ACCESSION AQ28386
 VERSION AQ28386.1 GI:3910204
 KEYWORDS GSS.

SOURCE
 ORGANISM human.
 HOMO sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 219)
 Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linhar,K., Golden,K.,
 Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
 Use of human BAC End Sequences for Sequence-Ready Map Building
 Unpublished (1998)
TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: madams@ligr.org
 Clones are derived from the human BAC library RPI1-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genetics (info@resgen.com). BAC end search page:
 http://www.tligf.org/tligf/humgen/bac_end_search/bac_end_search.html
 Seq primer: 77
 Class: BAC ends.

FEATURES source

1..219
 /organism="Homo sapiens"
 /db_xref="GDB:7529676"
 /db_xref="taxon:9606"
 /clone_lib="RPI1-11-78E13"
 /clone_lib="RPI1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPI11 Human Male BAC library"
 BASE COUNT 69 a 45 c 46 g 59 t
 ORIGIN

Query Match 5.1%; Score 157.2; DB 13; Length 219;
 Best Local Similarity 86.1%; Pred. No. 7.5e-31;
 Matches 174; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
 Qy 2099 tggctggaagcctcagttgtgctcctcagcactgtaagaacattatctcctcagtgag 2158
 |||||||
 Db 1 TGGCTGGAAGCCTTAGTGTGCTCTCAGACCTGAAGAACTTTATCTCTCATGCTGG 60
 |||||||
 Qy 2159 aagcagagcctcctcagcagatgagagagacacatcctgtaaacacccggaana 2218
 |||||||
 Db 61 AAGCCAGTCCTCTGACCAAGAGATGAGAGCACATCAATGTAACAAACCTGAAAA 120
 |||||||
 Qy 2219 ccttgatattcatgactacagaatcaacggtctccggtgtgctgactgacagcttgg 2278
 |||||||

Db 121 CCTGAGTATTCATGACCTACAGAAATCAAGAGGCTGCCGGTATTGTTATATCAATGCTG 180
 |||||||
 Qy 2279 gtaacttgaagaaccttacaac 2300
 |||
 Db 181 TGTGCTTGTTCACCTTAAAAA 202
 |||

RESULT 12
 BF903662 251 bp mRNA EST 18-JAN-2001
 LOCUS IL2-MT0180-181200-276-F03 MT0180 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF903662
 ACCESSION BF903662
 VERSION BF903662.1 GI:12295121
 KEYWORDS EST.
SOURCE
 ORGANISM human.
 HOMO sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 251)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?IL2-IL2&t2-IL2-MT0180-
 181200-276-F03&t3=2000-12-18&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 45
 High quality sequence stop: 96.

FEATURES source

1..251
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MT0180"
 /dev_stage="Adult"
 /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 ,716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 BASE COUNT 72 a 62 c 56 g 61 t
 ORIGIN

Query Match 5.0%; Score 153.8; DB 11; Length 251;
 Best Local Similarity 85.5%; Pred. No. 6.2e-30;
 Matches 218; Conservative 0; Mismatches 32; Indels 5; Gaps 4;
 Qy 1007 agaccctctcttgygcacacactgycacaatcagatgggt-gaaagtgaagtcacac 1065
 |||||||
 Db 1 AGACCCCTCTCTTTAGAGTCATCATTTGTGCAATCTCTATGACAGCAAAAGGAGTTCCAC 60
 |||||||
 Qy 1066 tctcacacaaacaaagcgtcttccatcctctatgatcgttgtatcacagaacaaacaa 1125
 |||||||
 Db 61 TCTCACACACAAACAGCTGATCATCTTATGATGATGAGATACAGAAAAACAAA 120
 |||||||

QY 1126 cacaacataaagtgctgcaagtgaactcgtgagagcctggaacctgtgagac 1185
 |||||
 Db 121 CACAACATTAAGGCGCTGTCTG--ATGTGACTCATTCGAGACCT-GACCACTGTGG-CAC 176
 |||||
 QY 1186 ctgagcttgagaggtgtgtctcccaagtttgatctggaacgagaggtgtccagc 1245
 |||||
 Db 177 CTGATCTGTGTGGTGCTGTCTGCGCACAAAGTTGAATTCGAATGCAAGATGTCCACCGC 236
 |||||
 QY 1246 gtgaatgagagatgctc 1260
 |||||
 Db 237 GCCTACTAGAAATGTC 251
 |||||
 RESULT 13
 LOCUS AO112439 630 bp DNA GSS 29-AUG-1998
 DEFINITION CIT-HSP-2372C1.TR CIT-HSP Homo sapiens genomic clone 2372C1, DNA
 sequence.
 ACCESSION AO112439
 VERSION AO112439.1 GI:3484599
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
 Berry,K., Grainger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
 Venter,J.C.
 TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
 Map Building
 JOURNAL Unpublished (1998)
 COMMENT Other-GSS: CIT-HSP-2372C1.TF
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.
 FEATURES
 Location/Qualifiers
 source 1..630
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2372C1"
 /clone_1id="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelobAC11. Site_1: HindIII. Site_2:
 HindIII"
 BASE COUNT 188 a 126 c 124 g 192 t
 ORIGIN
 Query Match 4.8%; Score 146; DB 13; Length 630;
 Best Local Similarity 100.0%; Pred. No. 8.5e-28;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2349 agctgaagcctgaaaacctggaagatgtgtttattcattgaccactgtctga 2408
 |||||
 Db 485 AGCTGAAGGCTCAAAAACCTGGAAGAAGATGTGTTTATTTCATTGACCCACTTGTCTGA 544
 |||||
 QY 2409 cattggaagaggaatgtagatatactcaagctctgttcaagtgaacctgtgaaccttga 2468
 |||||
 Db, 545 CATTGGAAGGAGGATGATACATAGTCAGAGCTCTGTCTCAAGTGAACCCGTGTGACCTTGA 604
 |||||
 QY 2469 agaataatcaatgtagtctcctgctgct 2494
 |||||

Db 605 AGAATTCATTAATGATCTCTCTGCT 630
 |||||
 RESULT 14
 LOCUS BF207840/c 840 bp mRNA EST 06-NOV-2000
 DEFINITION 601862546F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082074.5',
 mRNA sequence.
 ACCESSION BF207840
 VERSION BF207840
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
 plate: LICW44 row: 1 column: 11
 High quality sequence stop: 636.
 FEATURES
 Location/Qualifiers
 source 1..840
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4082074"
 /clone_1id="NIH_MGC_53"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
 Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGAGCGGCCGACATG-dT(30)BN-3'
 (where B = A, C, G, or T). Average
 insert size 1.55 kb (range 0.9-4.0 kb). 15/15 clones
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."
 BASE COUNT 276 a 143 c 165 g 256 t
 ORIGIN
 Query Match 4.6%; Score 139.8; DB 11; Length 840;
 Best Local Similarity 91.9%; Pred. No. 4e-26;
 Matches 203; Conservative 0; Mismatches 12; Indels 6; Gaps 5;
 QY 2853 tgatgtagctgccttcacatggtgattgagaacctgaagaatgattttttga 2912
 |||||
 Db 679 TGATGTAATGCTTCCTCCATAGGTGTA-TTGAAATCTTAACCAATTAGTGTT-TTTGA 622
 |||||
 QY 2913 cttagtactaagaatcttctacctgatacgaactagtcagaanaactgacc-aagtg 2971
 |||||
 Db 621 CTTTAGTACTTAAGAATTTCTACTGATCA-CATTAGTCTGAATACTTAGCCAAAGTGT 563
 |||||
 QY 2972 tatccaagtaactcttcgcaagaactagctgtgtgtggtgcaattgagatgatg 3031
 |||||
 Db 562 TATCCAAGTTAACTTTCTCAAGAAAGCTAGGCT-CCTTGAGTGTGATGATGATG 505
 |||||
 QY 3032 atccagatgattatcaagtggtctttaactagtaactgct 3072
 |||||
 Db 504 ATTCAGTGTATTACAGGTGCTTTTAACTAGTAAGTACTGCT 464
 |||||

RESULT 15
AM418826/c

LOCUS AM418826 509 bp mRNA EST 09-FEB-2000
DEFINITION ha21e11.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2874380 3',
mRNA sequence.

ACCESSION AM418826
VERSION AM418826.1 GI:6946758

KEYWORDS EST.

SOURCE human.

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdip/image/image.html
Seq primer: -400p from Glibco
High quality sequence stop: 461.
Location/Qualifiers

FEATURES
source

1. 509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2874380"
/clone_id="NCI_CGAP_Kid12"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 161 a 99 c 80 g 169 t
ORIGIN

Query Match

Best Local Similarity 4.4%; Score 136; DB 10; Length 509;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	2937	tgatccagcattagtcagaaaacttagccaggtgtatccaaagtaactttctcagaaga	2996
DB	509	tgatccagcattagtcagaaaacttagccaggtgtatccaaagtaactttctcagaaga	450
OY	2997	agctagcttcttggtgcaatttgatgatgatcatcgaatgtattacagtgctt	3056
DB	449	agctagcttcttggtgcaatttgatgatgatcatcgaatgtattacagtgctt	390
OY	3057	taactagtaactgct	3072
DB	389	TAAACTAGTAAGTCTCT	374

Search completed: March 25, 2002, 11:37:10
Job time: 2815 sec

THIS PAGE BLANK (USPTO)